

# A tRNA-mimic Strategy to Explore the Role of G34 of tRNAGly in Translation and Codon Frameshifting

A Janvier, L Despons, Laure Schaeffer, A Tidu, Franck Martin, Gilbert Eriani

▶ **To cite this version:**

A Janvier, L Despons, Laure Schaeffer, A Tidu, Franck Martin, et al.. A tRNA-mimic Strategy to Explore the Role of G34 of tRNAGly in Translation and Codon Frameshifting. International Journal of Molecular Sciences, 2019, 20 (16), pp.3911. 10.3390/ijms20163911 . hal-02294888

**HAL Id: hal-02294888**
**https://hal.science/hal-02294888**

Submitted on 28 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

International Journal of
*Molecular Sciences*

MDPI

*Article*

# A tRNA-mimic Strategy to Explore the Role of G34 of $tRNA^{Gly}$ in Translation and Codon Frameshifting

**Aurélie Janvier, Laurence Despons, Laure Schaeffer, Antonin Tidu, Franck Martin * and Gilbert Eriani ***

Architecture et Réactivité de l'ARN, Université de Strasbourg, Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire et Cellulaire, 2 allée Konrad Roentgen, 67084 Strasbourg, France
* Correspondence: f.martin@ibmc-cnrs.unistra.fr (F.M.); g.eriani@ibmc-cnrs.unistra.fr (G.E.)

Received: 4 July 2019; Accepted: 8 August 2019; Published: 11 August 2019




**Abstract:** Decoding of the 61 sense codons of the genetic code requires a variable number of tRNAs that establish codon-anticodon interactions. Thanks to the wobble base pairing at the third codon position, less than 61 different tRNA isoacceptors are needed to decode the whole set of codons. On the tRNA, a subtle distribution of nucleoside modifications shapes the anticodon loop structure and participates to accurate decoding and reading frame maintenance. Interestingly, although the 61 anticodons should exist in tRNAs, a strict absence of some tRNAs decoders is found in several codon families. For instance, in Eukaryotes, G34-containing tRNAs translating 3-, 4- and 6-codon boxes are absent. This includes tRNA specific for Ala, Arg, Ile, Leu, Pro, Ser, Thr, and Val. $tRNA^{Gly}$ is the only exception for which in the three kingdoms, a G34-containing tRNA exists to decode C3 and U3-ending codons. To understand why G34-$tRNA^{Gly}$ exists, we analysed at the genome wide level the codon distribution in codon +1 relative to the four GGN Gly codons. When considering codon GGU, a bias was found towards an unusual high usage of codons starting with a G whatever the amino acid at +1 codon. It is expected that GGU codons are decoded by G34-containing $tRNA^{Gly}$, decoding also GGC codons. Translation studies revealed that the presence of a G at the first position of the downstream codon reduces the +1 frameshift by stabilizing the G34•U3 wobble interaction. This result partially explains why G34-containing $tRNA^{Gly}$ exists in Eukaryotes whereas all the other G34-containing tRNAs for multiple codon boxes are absent.

**Keywords:** genetic code; tRNA; glycine codon; translation; IRES element; frameshifting

## 1. Introduction

The process of translation is the last stage in the genetic information transfer and it depends upon the correct matching between mRNA codons and corresponding tRNA anticodons within the ribosomal complex. High fidelity of tRNA selection is crucial in the expression of genetic information. On the ribosome, each codon of the mRNA has to be accurately and efficiently recognized by cognate tRNAs among a large repertoire of non-cognate tRNAs. Aminoacylated tRNAs (aa-tRNAs) are delivered to the ribosomal A site as a ternary complex with elongation factor EF-Tu and GTP. Incorrect tRNA can be rejected either before or after GTP hydrolysis. During the initial selection step, the ternary complex of aa-tRNA with EF-Tu–GTP (in procaryotes) binds to the ribosome and forms a readily reversible complex that dissociates rapidly when there is no codon/anticodon match. This step occurs without GTP hydrolysis. Most non-cognate tRNAs are rejected by this first sieve with essentially no cost with respect to GTP hydrolysis [1]. Discrimination of near-cognate ternary complexes is enhanced by a proofreading step that requires a more complex mechanism with the hydrolysis of GTP in EF-Tu. After GTP hydrolysis, EF-Tu undergoes a dramatic conformation change, thereby loosing its affinity for aa-tRNA. The aminoacyl end of aa-tRNA is thus free to move into the peptidyl transferase centre

on the 50S subunit (accommodation), where it takes part in peptide bond formation. Alternatively, after GTP hydrolysis, near-cognate aa-tRNA with imperfect base pairing dissociate from the ribosome. Overall, discrimination of correct and incorrect substrates is achieved in two consecutive steps, initial selection and proofreading which are separated by GTP hydrolysis. All these events between the initial binding and accommodation of the cognate aa-tRNA have been dissected into several distinguishable steps (reviewed in [2,3]).

Although the process of tRNA selection on the ribosome seems to be universal, examples of decoding of codon-anticodon pairs that deviate from the canonical process have been described. Dicistroviruses such as cricket paralysis virus (CrPV) contain between their two ORFs, an IRES element called intergenic region (IGR) that mimics the structure of tRNA presenting to the ribosome a pseudo codon-anticodon triplet. The IGR forms a pseudo knot (PKI), which mimics the codon-anticodon nucleotides. It enters the A-site, translocates to the P-site, and initiates translation on the next codon, a non-AUG initiation codon [4,5]. Translation initiation is independent from the sequence of the triplet present in the PKI and can occur with a stop codon [5]. However, when the pseudo codon-anticodon triplet is disrupted by a mutation in the PKI sequence, the IGR is unable to initiate translation on the downstream ORF [5]. In short, the IGR mediates ribosome hijacking in achieving a sophisticated molecular mimicry of tRNA and mRNA. For that, the IGR adopts a modular structure composed of 6 stem-loops and 3 pseudo knots, balanced between rigidity and flexibility in order to optimally interact with the ribosome and elongation factor eEF2 during translocation [6–9]. It was shown that PKI is recognized by the ribosome in the same way as a cognate mRNA-tRNA codon-anticodon duplex. Stem loops IV and V of the IGR play a significant role during this process by interacting with the head of the ribosome to restrict its movement as observed during canonical translation [9–11]. Once PKI has entered the P-site, the contacts with stem loops IV and V are disrupted, leaving the A-site free to accept the first aminoacylated tRNA complexed with elongation factor eEF1A [12]. After a second translocation, the IGR mimics a canonical acceptor stem of an E-site tRNA [13]. Altogether, the data show that PKI of the IGR follows the conventional molecular mechanisms of tRNA discrimination by sensing the codon/anticodon duplex in the A site and the subsequent step of translocation in the P site [9]. Following 80S assembly, the eEF2-mediated translocation of the IGR in the P-site results in an unstable intermediate that is captured by binding of the elongator tRNA in complex with eEF1A. At this step, initiation can occur in the main frame (0 frame) and also after frameshifting in the +1 frame, according to the arrival rate of the first tRNA [14]. Usually, frameshifting during decoding generates much more drastic effects than missense errors that change a single amino acid residue (estimated to be in the range $6 \times 10^{-4}$–$5 \times 10^{-3}$) [15]. It occurs when the ribosome recruits an aa-tRNA out of frame with respect to the frame in which it was initiated. The most common of these events require the ribosome to shift to a codon overlapping a codon in the normal frame. A shift of one nucleotide toward the 5′ end of the mRNA generates a (−1) frameshifting, whereas a shift toward the 3′ end generates a (+1) frameshifting. Shifted ribosomes that cannot finish the normal protein. They no longer see the correct codons, and normally rapidly encounter a termination codon. The importance on the protein function is usually consequent because of the sequence change and premature stop. Although the spontaneous frequency of frameshift errors is below missense errors, it can be considerably increased by additional factors such as slippery sequences or ribosomal pausing on hungry codons that modify the codon-anticodon realignment [16]. In addition, some frameshifts are programmed by primary and secondary sequence elements either preceding or following the shift site [16–18].

Multiple types of molecules in the translational machinery can promote translational frameshifting, including ribosomal RNA, ribosomal proteins, and translation elongation factors [16]. A common molecule affecting the frame maintenance is the tRNA. Frameshift suppressor tRNAs often harbour alterations of the loop of the anticodon but also normal 7-nt anticodon loops [19]. Changes in the tRNA base modifications such as N1-methylguanosine at position 37, are also able to induce frameshifting [20–22]. Recently, a structural study revealed that suppressor tRNASufA6 ($tRNA^{Pro}$ with an expanded anticodon loop) is undergoing a rearrangement of the anticodon loop due to

the extra nt at position 37.5 that destabilizes U32, thereby disrupting the conserved U32-A38 base pair [23]. Interestingly, the removal of the $m_1G37$ modification of $tRNA^{Pro}$ disrupted U32-A38 pairing in a structurally analogous manner, explaining the frameshifting effects previously observed [24]. Altogether, these data showed that the disruption of the U32–A38 base pair plays a significant role in the loss of grip in the correct frame. Phylogenetic analysis has shown that the 32–38 interaction is a fundamental feature of all tRNAs and is directly correlated to the strength of the codon–anticodon interaction (high vs. low GC base pairs) [20,21]. Modifying the 32–38 interaction dramatically affects tRNA discrimination of cognate vs. near-cognate codons both in vitro and in vivo [25,26].

Much of what we know about frameshifting induced by tRNAs has come from studies on suppressors of $tRNA^{Gly}_{NCC}$ and $tRNA^{Pro}_{NGG}$. Sequencing of suppressible frameshift mutations showed that they include the sequences CCC-N or GGG-N, resulting from expansion of proline (CCN) or glycine (GGN) codons. This allows the tRNA suppressor to frameshift from the original phase 0 to the shifted phase (+1) while keeping the codon-anticodon pairing. Glycine belongs to the GC-rich 4 codon boxes that establish strong base pairing with their corresponding anticodon. The structures of these tRNAs have been 'tuned down' during evolution to reduce and adjust their binding strength to the mRNA in the ribosome [27,28]. The adjustment structure is obtained by varying nucleotides forming the interaction 32–38 in the anticodon loop. These nucleotides should not form a Watson–Crick base pair or another interaction that would antagonize the canonical conformation of the codon/anticodon triplet in a correct helical orientation in the ribosomal decoding site. Without a selection for the preformed conformation of the anticodon loop, the strong codon/anticodon pairs would allow miscoding by binding to other G/C-rich near- or even non-cognate aatRNAs.

In eukaryotes, three $tRNA^{Gly}$ isoacceptors are found with GCC, CCC and UCC anticodons. According to the wobble rules, $tRNA^{Gly}_{UCC}$ should be able to decode GGA and GGG codons. The $tRNA^{Gly}_{GCC}$ isoacceptor should decode GGC and GGU codons. Theoretically, these two tRNA species, that are the most abundant, should be able to decode the entire set of glycine codons, however, an additional $tRNA^{Gly}$ with CCC anticodon is found to decode GGG codons [29]. Nucleotide U34 of $tRNA^{Gly}_{UCC}$ appears to be unmodified in higher eukaryotes, suggesting that it may also decode all 4 Gly codons by "superwobbling" [30]. Remarkably among eukaryotic tRNAs reading the 4 and 6 codon boxes, $tRNA^{Gly}$ is the only one that exhibits a G34. Other tRNAs have a A34 that is modified in inosine to decode NNU, NNC and NNA codons (the later inefficiently [31]). Therefore, the decoding ability of G34 and I34 does not drastically differ and the reason of this dichotomy is not yet understood. Recently, an interesting correlation was described between the A34->I modification and the proofreading capacity of corresponding aminoacyl-tRNA synthetases (aaRS), the enzymes in charge of tRNA aminoacylation [32]. It was suggested that the proofreading activity would prevent continued sectoring of the genetic code utilizing the anticodon wobble position. The A34->I modification, found in tRNAs reading 4-codon boxes, would block its subdivision in two 2-codon boxes and thereby impact the evolution and sectoring of the genetic code [32]. Gly would escape to this trend, perhaps because glycyl-tRNA synthetase is highly specific for Gly and does not required proofreading activity. Therefore, Gly would not be concerned by codon reattribution and $tRNA^{Gly}$ could use a G34 wobble and act as a '2-codon box reader' without the risk of inducing the split of its 4-codon box [32]. In short, it seems that $tRNAs^{Gly}$ bring together several features and properties that make them singular. They decode GC-rich codons, implying an adaptation of the anticodon loop structure to balance the energy of the codon-anticodon helix. They do not exhibit modification of A37, which is usually involved in reading frame maintenance. Many suppressors of $tRNA^{Gly}$ that frameshift in +1 have been be selected [33] (and ref inside) and programmed ribosomal frameshifting on GGG codon in −1 frame is utilised by HIV-1 virus to modulate GagPol synthesis [34]. All these features suggest that $tRNAs^{Gly}$ may play a general and significant role in the phenomenon of programmed ribosomal frameshifting, which regulatory role in both prokaryotes and eukaryotes is becoming increasingly clear [35–38] (and references therein).

In this context and to further explore the decoding and frameshifting properties of $tRNA^{Gly}$ we performed the present study which combines experimental and bioinformatics approaches to carry out a systematic analysis of the Glycine codon usage in Eukaryotes and relationship with the 3′ adjacent nucleotide (nt+1). We found a clear distribution preference for a G over the 3 other nucleotides immediately after the GGU codon, from which we infer that this nucleotide likely causes an effect on translation. We assessed the effect of the four nucleotides at n+1 position in rabbit reticulocyte extracts by monitoring renilla luciferase activity of constructs programmed with the tetranucleotide GGU(N) in the two frames 0 and +1. To control both codon and anticodon pairings we used the internal ribosome entry site (IRES) from cricket paralysis virus (CrPV) that initiates using a pseudoknot structure mimicking a tRNA paired with the mRNA. We found that nt+1 regulates the frame selection by the IRES. Moreover, a G at +1 improves luciferase activity and favours the frame 0. Altogether, the results suggest that translation is modulated by the first nucleotide of the adjacent codon within the ribosome.

## 2. Results and Discussion

### 2.1. $tRNAs^{Gly}$ Have Unusual Nucleotide 34 Distribution

The current 'genomic tRNA data base' contains in the eukaryota domain 139,808 tRNA genes (http://gtrnadb.ucsc.edu/) [29]. Analysis of these data confirmed previous observations on the poor usage of $G_{34}NN$ anticodons by tRNAs specific of amino acids coded by 3, 4 and 6-codon boxes. Only 3063 sequences have been reported for the following amino acids: Ala, Arg, Ile, Leu, Pro, Ser, Thr and Val, which represents less than 4% of the total number of tRNA sequences (Table 1). Remarkably, $tRNA^{Gly}$, which also belongs to the 4-codon box group, clearly escapes this global trend and 6636 sequences harbouring an anticodon $G_{34}CC$ are found in the database. In total, 13,347 sequences of $tRNA^{Gly}$ have been reported in eukaryotes and the anticodon distribution is 767 for ACC, 2328 for CCC and 3616 for UCC in addition to the 6636 for GCC. Another surprising aspect of this distribution is the low number of genes for $tRNA^{Gly}$ with ACC anticodons (only 767) compared to the other tRNA specific of the 3, 4 and 6 codon boxes. For instance there are 40,726 sequences exhibiting A34 in tRNAs specific for Ala, Arg, Ile, Leu, Pro, Ser, Thr and Val and only 767 sequences of $tRNA^{Gly}$. It results a spectacular inversion of the usage of G34 versus A34 when comparing $tRNA^{Gly}$ numbers with the other tRNAs specific for Ala, Arg, Ile, Leu, Pro, Ser, Thr and Val (Figure 1). Altogether, the data are suggesting that $tRNA^{Gly}$ prefers G34 rather than A34 that could be potentially modified in I34.

**Table 1.** Identity of nucleotide 34 in tRNAs in the 3, 4 and 6 codon boxes *.

| tRNAs Specific for | Total Number in Eukarya [1] | With G34 | With A34 | With C34 | With U34 |
|---|---|---|---|---|---|
| Ala, Arg, Ile, Leu, Pro, Ser, Thr and Val | 80,060 | **3063** | **40,726** | 14,194 | 22,077 |
| Gly | 13,347 | **6636** | **767** | 2328 | 3616 |

[1] from http://gtrnadb.ucsc.edu/ [29]; * Excluding in the 6-codon boxes the tRNAs for the 2-codon boxes.

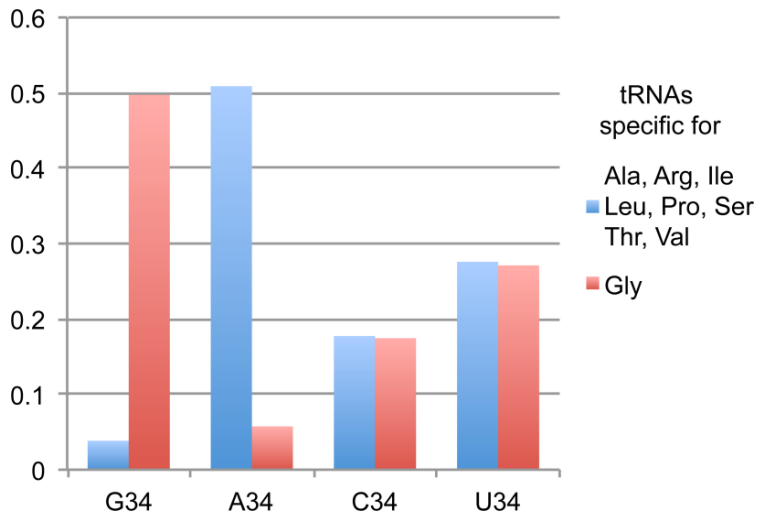



**Figure 1.** Distribution of nucleotide 34 in tRNAs specific for 3, 4 and 6 codon boxes according to the amino acceptor identity. The bars represent the usage of each nucleotide 34 in the subgroups containing $tRNA^{Gly}$ alone or tRNAs specific for Ala, Arg, Ile, Leu, Pro, Ser, Thr and Val. G34 is overrepresented in $tRNA^{Gly}$ and nearly absent in tRNAs specific for Ala, Arg, Ile, Leu, Pro, Ser, Thr and Val.

### 2.2. Usage of Codons Gly in Eukaryotes Does Not Show Any Bias

Next we checked whether sparing A34 in $tRNA^{Gly}$ could result from a special codon usage. We examined the 'Codon Usage Database' of a list of 23 representative eukaryotic organisms including *Candida glabrata, Candida albicans, Saccharomyces cerevisiae, Kluyveromyces lactis, Drosophila melanogaster, Schizosaccharomyces pombe, Leishmania major, Anopheles gambiae, Oryza savita, Zea mays, Arabidopsis thaliana, Papio anubis, Sus scrofa, Strongylocentrotus purpuratus, Rattus norvegicus, Homo sapiens, Mus musculus, Macaca mulatta, Equus caballus, Oryctolagus cuniculus, Xenopus tropicalis, Gallus gallus, Caenorhabditis elegans* [39]. Despite variations among the organisms, we did not observe any special bias in the usage of Gly codons that would explain the presence of G34 (Figure 2).

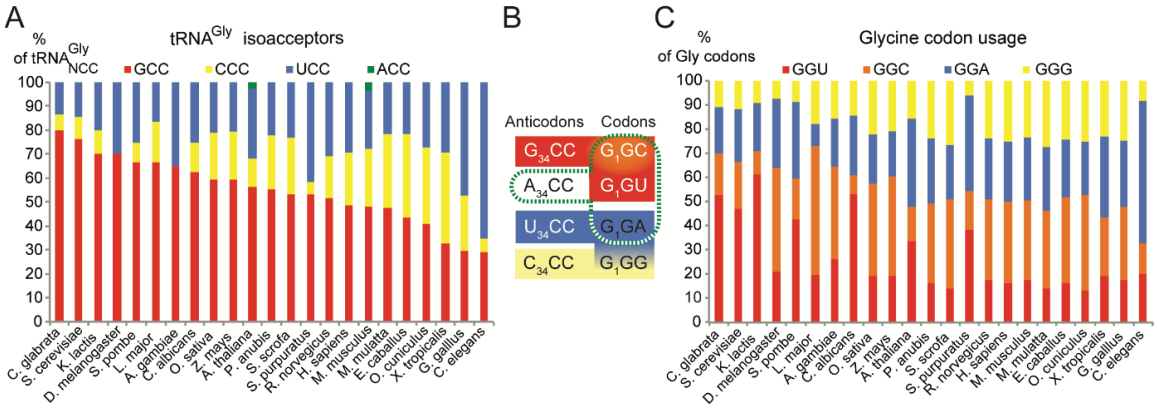



**Figure 2.** Distribution of $tRNA^{Gly}$ isoacceptors and Gly codon usage in a representative set of eukaryotes. (**A**) Histogram representing the distribution of $tRNA^{Gly}$ isoacceptor genes per organism expressed as percentages (data from http://gtrnadb.ucsc.edu/). (**B**) The codon reading properties of $tRNAs^{Gly}$ are represented with boxes using the colour code used in the histograms of panels A and C. The decoding properties of the rarely found $tRNA^{Gly}_{A34CC}$ (modified in $I_{34}CC$) are represented with a dashed green box. (**C**) The histogram represents the codon usage for Gly codons per organism, expressed as percentages (data from http://www.kazusa.or.jp/codon/).

As mentioned above, three main $tRNA^{Gly}$ isoacceptors are found in eukaryotes, with GCC, CCC and UCC anticodons. According to the wobble rules, $tRNA^{Gly}_{UCC}$ should decode GGA and GGG

codons whereas $tRNA^{Gly}_{GCC}$ isoacceptor should decode GGC and GGU codons. A third $tRNA^{Gly}$ with CCC anticodon is found to specifically decode GGG codons [29] (Figure 2B). Therefore, the decoding of Gly codons by $tRNA^{Gly}$ is perfectly ensured despite the absence of an A34-containing $tRNA^{Gly}$. It is assumed that such tRNA, if existing, would be deaminated on A34 to form inosine with C, U and A pairing properties (however inefficient with A [31]). Inosine has a greater range of potential base pairs than either adenosine or guanosine. However, I•C and I•U pairs are less stable than G•C and G•U pairs, respectively [40]. In eukaryotes, the G34 anticodon-sparing has been observed for a long time with the exceptional disparity of $tRNA^{Gly}$ [41,42]. In the three kingdoms, for an unexplained reason, G34-containing $tRNA^{Gly}$ read the pyrimidine-ending codons (GGU and GGC) and favor G•U wobble versus I•A, I•C or I•U wobble pairs that would be formed if A34-$tRNA^{Gly}$ modified in inosine would exist. Altogether, these observations suggest that the preference given by $tRNA^{Gly}$ for G34 versus A/I34 cannot be explained by decoding idiosyncrasies of the Gly codons or by economical reasons in terms of tRNA anticodon number.

### *2.3. Bias of Nucleotide +1 after GGU Glycine Codon*

One hypothesis might be that G34 is part of a mechanism of adaptation of $tRNA^{Gly}$ linked to the inherent properties of frameshifting of this tRNA. Many records have described the frameshifting properties of suppressors of $tRNA^{Gly}_{NCC}$. They generally frameshift on sequences GGG-N, resulting from expansion of glycine (GGN) codons.

Therefore, we examined the contexts of Gly codons, focusing on nt+1 that could be involved in +1 frameshifting phenomenon. We analysed all the open reading frames (ORF) of the 23 above-mentioned organisms for their distribution in Gly codons and +1 context. Data processing was performed using a home-written Python script. Coding DNA sequences (CDSs) from genomic sequences of 23 eukaryotic species were downloaded from the NCBI site ftp://ftp.ncbi.nlm.nih.gov/genomes/. 35,246,252 Gly codons were found in the 23 annotated genomes. The histogram of Figure 3 shows the composition of the nt+1 after the four Gly codons, in all the ORFs of the 23 organisms. In the GGU box a clear preference was observed for a G at +1 in 42% of the codons (Figure 3). Normally it is expected to find an equal ratio for the 4 nts with little variation according to the G/C content of the genome. Here, the G preference is specific of the GGU codon and reaches an unusual high value. Although some discrepancies for nt+1 also exist in the other 3 Gly codons, they are less significant (Figure 3). The distribution was examined in each separate organism and the same general pattern of G preference in +1 position of the GGU codons was found with extreme values of 59.7% in *L. major* and 34% in *S. pombe* (Figure A1 in Appendix A).

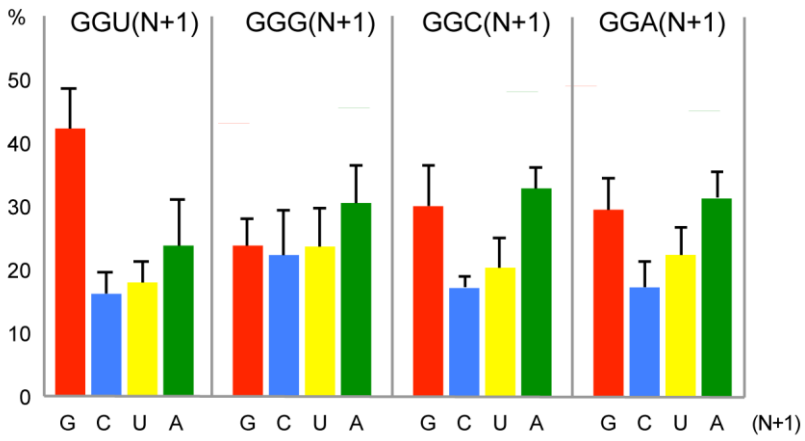



**Figure 3.** Distribution of nucleotide +1 downstream from the four Gly codons in 23 eukaryotic genomes. 35,246,252 Gly codons were analysed. Results are expressed as percentages. The standard deviation bars are showing the dispersion between the genomes. The *x*-axis corresponds to the nucleotides found at position +1 after the four Gly codons.

We also examined the $tRNA^{Gly}$ gene number. The fraction of $tRNA^{Gly}_{GCC}$, which decodes GGU codons, varied between 0.29 and 0.80 (Figure 2C) and was well adapted to the Gly codon usage (Figure 2A).

### 2.4. Decoding and Frameshifting Efficiency on Gly Codons

To measure the frame shifting activity on Gly codons we used the second IRES element of CrPV located in the intergenic region (IGR) that initiates using a pseudo knot structure mimicking a tRNA paired with the mRNA. Like the transfer-messenger RNA (tmRNA), the IGR combines features of tRNA and mRNA [43]. However, while tmRNA is involved in the degradation of aberrant mRNAs, the IGR structure is dedicated to translation initiation of the second cistron of the CrPV. A pseudoknot (PKI) of the IGR structurally and functionally mimics the codon-anticodon helix and is recognized by the ribosome in the same way as a cognate mRNA-tRNA codon-anticodon duplex [44]. While tRNAs have low affinity for ribosome in the free form, the IGR binds to the A-site independently of the elongation factor. Then, the pseudo-tRNA is translocated in the P-site, and translation is initiated on the next codon, a GCU in CrPV [6–8]. During this process on the ribosome, the IGR undergoes a careful monitoring of its structure, which must perfectly fit with the geometry of a cognate codon/anticodon triplet as found in a tRNA-mRNA complex. For instance, a single mutation that disrupted the pseudo codon-anticodon triplet was shown not to have dramatic consequence on viral protein synthesis. On the other hand, the IGR was perfectly able to start with a stop codon inserted in the pseudo knot [4,5]. Here in this study, the IGR was used as a versatile tool for testing the frameshifting activity induced by the environment of Gly codons. The pseudoknot triplet was reprogrammed with Gly codons and fused with renilla luciferase (R-Luc) ORF in order to initiate in the 0 phase. The insertion of one extra A before the AUG initiation codon of R-Luc induced a +1 frameshift that inactivated R-Luc translation. This construct was used to measure R-Luc activity resulting from the +1 frameshift (+1 frame). By this way we could reprogram the codon and anticodon parts of the IGR in order to explore their reading properties on the ribosome decoding center, and measure the influence of nt+1 on the reading frame maintenance with the R-Luc constructs in 0 and +1 frames.

#### 2.4.1. Decoding and Frameshifting Efficiency of the WT IGR

In a first series of experiment, we tested the activity of the IGR with a wild-type PKI with a G in position +1 of the first codon of R-Luc (G+1 hereafter) and compared it with derivatives with the 3 other nucleotides in +1 (A+1, C+1 and U+1). Transcripts of the constructs in 0 and +1 frame of R-Luc were translated in rabbit reticulocyte lysate (RRL) and the resulting luciferase activities were quantified.

We first measured the R-Luc activity of the wild-type (WT) PKI with the complementary sequence AGG/CCU mimicking a $tRNA^{Pro}_{AGG}$ annealed to a CCU Pro codon. In parallel, a negative IGR control was constructed with a disrupted PKI exhibiting a GGU codon instead of CCU [5]. The disruption strongly diminished the 0 and +1 frame translation activity to a level of 4% of the activity of the WT IGR in 0-frame (Figure 4). The background was identical in 0 and +1 frame, and could account from the spontaneous cap-independent translation activity typically measured with RRL extracts. This background was subtracted from the measured activities in all the following experiments.

With the WT IGR, a spontaneous frameshifting activity was observed when comparing the activity of R-Luc in 0 and +1 frame. This is consistent with other studies performed on the IGR of other dicistroviruses that showed translation of an overlapping ORFx in the +1 frame [44]. The R-Luc activity in +1 frame reached 7% of that of 0 frame, quite similar to the activity previously reported [44–46]. We also observed that the ratio of synthesis of the two frames is sensitive to the $Mg^{2+}$ concentration and translation conditions as described [47]. Optimal $Mg^{2+}$ concentration to add was found to be 1 mM in our activated RRL translation extracts. With 0.5 mM or 2 mM $Mg^{2+}$ the extracts were less active and less specific for frame selection. They were entirely inactive above 2 mM (Figure A2).

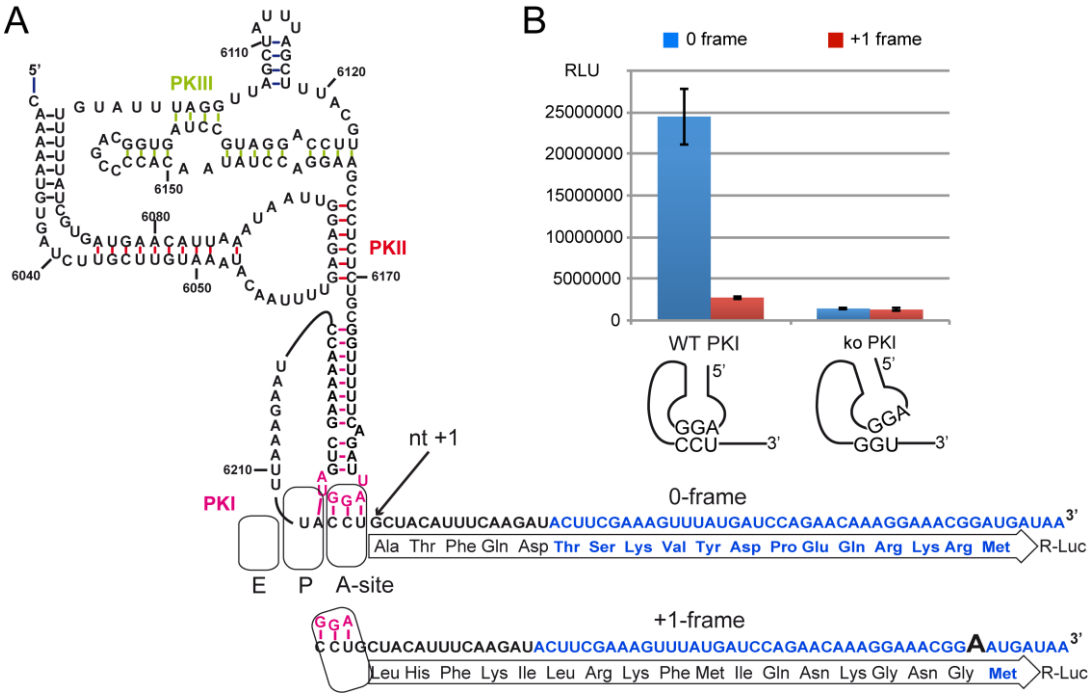



**Figure 4.** Translation of R-Luc driven by the CrPV IGR and analysis of potential +1-frame translation. (**A**) The secondary structure of the IGR IRES. The IGR is driven translation of Renilla luciferase (R-Luc) fused in the 0-frame or +1-frame. To monitor the +1-frameshifting activity, one extra A nucleotide was added in the R-Luc sequence. Nucleotides and amino acid residues in blue are from R-Luc. The pseudoknot PKI which mimics the codon/anticodon is drawn in pink. The A-, P- and E-site of the ribosome are schematized. Nucleotide numbering is from the CrPV genome. (**B**) Histogram representing the luciferase activities in 0-frame or +1-frame starting from the WT or knockout PKI. Transcripts were synthesized in vitro, denatured/renatured and incubated in Rabbit Reticulocyte Lysates at 30 °C for 2 h. Translation of R-Luc was measured by measuring the bioluminescence produced with a Dual-Glo luciferase assay (Promega) in a luminometer. Translation efficiencies are represented as raw bioluminescence activity (Relative Light Units or RLU) and resulted from at least 3 independent experiments. Standard deviations are shown.

Then we analysed the impact on translation efficiency of the substitutions of the first nt after the WT CCU Pro codon of PKI. These mutations change the first amino acid and may potentially affect translation initiation efficiency and frame selection. Dicistroviruses are able to initiate with different amino acids. Plautia stali intestine virus (PSIV) starts with Gln (CAA), Israeli acute paralysis virus (IAPV) with Gly (GGC) and many other dicistroviruses start with Ala (GCU/C/A) [48]. Here we generated the 3 constructs of the WT PKI with an A, C or U in +1. The resulting new codons were ACU, CCU and UCU leading to Thr, Pro and Ser changes, respectively. These 3 amino acids are considered to be non-destabilizing residues according to the N-end rule and therefore should not affect differently the half-life of the translated products [49]. However, when comparing the R-Luc activities, we could observe significant variations with the 4 constructs. The most active construct in 0 frame was the WT IGR with G in +1 downstream of non-coding Pro codon of PKI. This result suggests that evolution has already selected the most efficient amino acid to initiate de synthesis of the viral protein. Compared to the WT PKI with a G+1, the other constructs with A+1, C+1 and U+1 exhibited 39%, 15% and 13% of activity, respectively (Figure 5). Therefore the nt in +1 significantly impacts translation initiation despite its location outside of PKI. In the cryo-EM structure of the CrPV IGR in complex with *K. lactis* 80S, an additional base pair is observed between G+1 and the ribosomal base C1273 that is usually interacting with the third base pair of the anticodon during standard decoding [9]. Remarkably, all IGR-IRES sequences obtained to date encode a guanine residue at this position with the only exception

being the PSIV-IRES [48]. This conservation correlates with the function of G+1 in mRNA frame selection [47], and the variations of translation efficiency here observed with the 3 variants.

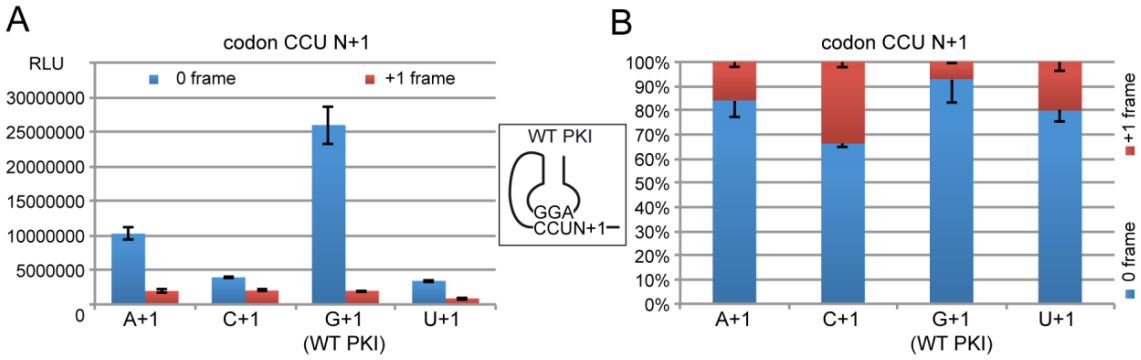



**Figure 5.** The effect on translation of the first nucleotide downstream the wild-type pseudo knot (PKI). (**A**) Histogram representing the luciferase activities of the synthesized R-Luc in 0-frame or +1-frame. Translation efficiencies are represented as raw bioluminescence activity (Relative Light Units or RLU) and resulted from at least 3 independent experiments. Standard deviations are shown. The construct with a G+1 corresponds to the WT PKI. (**B**) The same data represented as a 100% stack bar graph.

Interestingly, the consequent decreases of activity in 0 frame were not correlated with increases of activity in the +1 frame that would suggest competition for ribosome site selection in the 0 or +1 frame (Figure 5). However, the ratio of R-Luc activity in +1 and 0 frame showed significant variations from 0.07 for the WT IGR to 0.19, 0.51 and 0.25 for the variants with A+1, C+1 and U+1, respectively. These data confirmed that nt+1 is involved in frame selection [47] although the loss of activity in 0 frame was not balanced by an increase of activity in the +1 frame. The molecular mechanism of this frameshifting is not yet clearly understood, but it has been proposed that on the IGR, the ribosome is oscillating between 0 and +1 frames and the frame is locked when the cognate tRNA binds [14]. At this step, the guanine residue in +1 might play a critical role in stabilizing the conformation more suitable for the 0 frame. Therefore, the tRNA arrival rate would dictate the choice of the 0 and +1 frame to translate [14]. This is consistent with the fact that $tRNA^{Ala}_{AGC}$ for the 0 frame is an abundant isoacceptor whereas $tRNA^{Leu}_{UAG}$ for the CUA codon of +1 frame is a relatively rare isoacceptor [29].

### 2.4.2. Decoding Efficiency of the IGR Programmed with GGU Gly Codon

First, we tested PKI constructs programmed with a GGU Gly codon. As mentioned earlier, GGU codon is decoded by $tRNA^{Gly}_{GCC}$. We replaced the CCU codon of the WT PKI by the GGU Gly codon and to preserve the complementarity with the pseudo-anticodon of PKI, we inserted the GCC sequence in the opposite strand (Figure 6). The resulting codon-anticodon minihelix contained an equivalent 'G34•U3' wobble pairing, a combination never tested in the IGR. Indeed, previous mutagenesis studies established the essential role of the correct base pairing of the tRNA/mRNA mimicry in PKI of the IGR [4,5]. As the IGR mimics a tRNA paired with the mRNA, it is possible to test all the codon/anticodon combinations of interest, including non-canonical base pairing.

As before, the reprogrammed PKI was tested with the 4 nts in position +1 and luciferase activity was assessed in 0 and +1 frame (Figure 6). Of the 4 constructs, GCC/GGU-G was the most efficient with 74% of the R-Luc activity measured with the wild-type AGG/CCU-G construct. Therefore a construct with 'G34•U3' wobble pair is well recognized by the ribosome in the presence of a G in +1. The 3 other constructs exhibited severe decreases of the activity in 0 frame with 21, 34 and 11% of WT activity for A, C and U in +1, respectively (Figure 6). These results confirmed the critical role of nt+1 for the IGR activity even with a PK reprogrammed with a Gly codon.

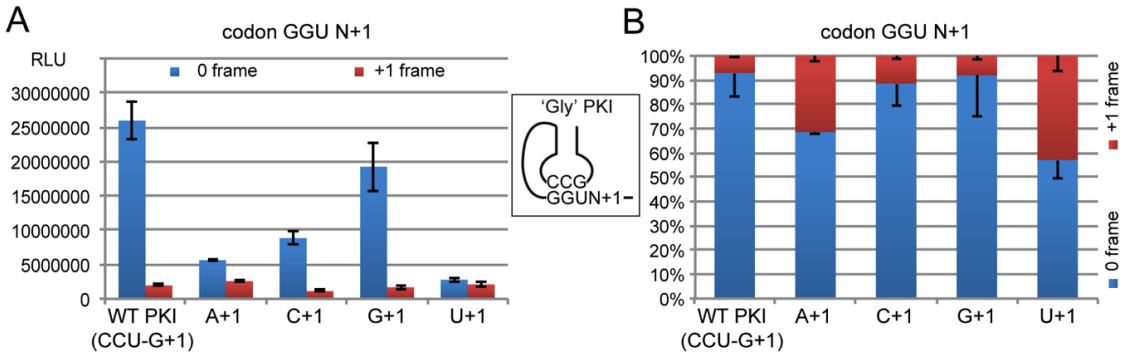



**Figure 6.** The impact on translation of the first nucleotide downstream of the PKI reprogrammed with a Gly GGU codon. (**A**) Histogram representing the luciferase activities of the synthesized R-Luc in 0-frame or +1-frame. Translation efficiencies are represented as raw bioluminescence activity (Relative Light Units or RLU) and resulted from at least 3 independent experiments. Standard deviations are shown. The first construct corresponds to the WT PKI. (**B**) The same data represented as a 100% stack bar graph.

### 2.4.3. Decoding Efficiency of the IGR Programmed with Other Gly Codons

In addition to the IGR programmed with the GGU codon, we constructed variants with the other Gly codons. A first variant contained a GGU Gly codon (with G+1) annealed to an ACC anticodon instead of GCC. The (ACC/GGU-G) variant mimics recognition by the absent $tRNA^{Gly}_{A34CC}$, which in eukaryotes is specifically replaced by the anticodon with a G34. Here a regular Watson-Crick pairing 'A34-U3' is expected to form in the ACC/GGU-G variant instead of the 'G34•U3' wobble pairing found in the GCC/GGU-G construct. The RNA with the Watson-Crick 'A34-U3' pair was nearly as active as the construct with 'G34•U3' wobble pair (55% and 65% of the activity of the WT IGR, respectively) (Figure 7). Thus, the absence of $tRNA^{Gly}_{A34CC}$ is not due to a fundamental structural or functional defect hampering its function in translation. Nevertheless, the RNA with 'A34-U3' pair is 10% less active than the 'G34•U3' whereas the frameshifting activity is increased to 9% (versus 7% for the 'G34•U3' construct). Therefore, the GCC anticodon forming a wobble 'G34•U3' with the GGU codon is the most efficient and faithful combination.

Four other constructs were generated to mimic the other codon-anticodon combinations formed during codon Gly decoding. Transcripts GCC/GGC, CCC/GGG, UCC/GGA, and UCC/GGG (all with G in +1) were constructed in order to test initiation on GGC, GGG and GGA codons in the PK, respectively. The 3 first transcripts exhibited 35% of residual activity in 0-frame translation whereas the fourth one, the UCC/GGG variant, had lost nearly all its activity in frame 0, which was below the level of the +1 frame activity. The low activity of the UCC/GGG variant with a 'U34•G3' wobble pair may be surprising since the GCC/GGU variant with the reverse 'G34•U3' pair exhibited 65% of the WT activity. This result confirms that G•U and U•G pairs are not isosteric [50] and confer a non-functional geometry to the IGR in the 'U34•G3' orientation. Usually U34-containg tRNAs utilize diverse modifications to stabilize tautomer conformations suitable for the ribosomal A-site [51]. However here, the transcripts of the IGR are devoid of any modification and therefore the PKI 'anticodon loop' cannot achieve the suitable conformation change. This might explain the drastic effect on the activity of the 0-frame while the frame-shifting activity was kept nearly constant.

Overall, the data collected with the 5 constructs are showing that the presence of two strong G=C pairs in codon positions 1 and 2 do not guarantee the full activity to the IGR. The more active constructs are exhibiting a weak third pair such as a wobble pair in the 'G34•U3' orientation or an A-U pair. Three G=C pairs are reducing the activity as well as the reverse wobble pair 'U34•G3'.

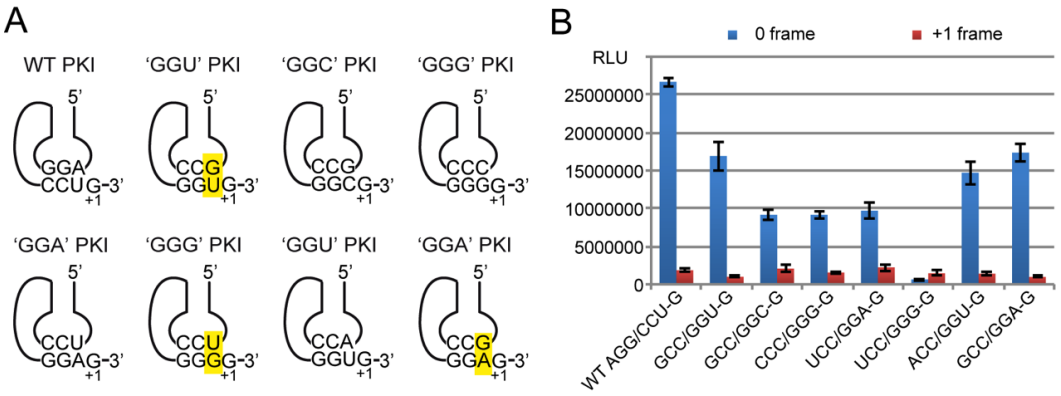



**Figure 7.** The effect on translation of reprogramming the PKI with different Gly codon-anticodon combinations. (**A**) Schematic representations of the different RNAs tested. Non Watson-Crick base pairs are highlighted in yellow. (**B**) Histogram representing the luciferase activities of the synthesized R-Luc in 0-frame or +1-frame. Translation efficiencies are represented as raw bioluminescence activity (Relative Light Units or RLU) and resulted from at least 3 independent experiments. Standard deviations are shown. Construct names are described as XXX/YYY-G where X represent anticodon triplets, Y codon triplets, all with a G in +1 and written in the 5′-3′ direction. The first construct corresponds to the WT PKI.

To confirm this observation, an additional construct GCC-GGA with a non-Watson-Crick pair at position 'G34•A3' was constructed and tested. The construct was as active as the best Gly construct (Figure 7). G•A pairs are supposed to bring flexibility to RNA structures [52–55]. Up to six dominant modes of G•A pairing have been reported in RNA molecules comprising ribosomes, riboswitches, ribozymes, tRNAs, mRNAs, etc. [56] and the geometry of a G•A pair was solved at the central position of the codon-anticodon helix in the 70S ribosome [57].

Altogether, the data show that in the Gly codon context, the presence of more than two G=C Watson-Crick pairs in the PK is deleterious for the translation initiation activity. The presence of A-U at the third codon position is tolerated, but a weaker G•U wobble pair or a G•A non-Watson Crick pair is conferring the highest and faithfulness translational activity to the IGR. All these results are consistent with our knowledge about the ribosomal A-site. It recognizes precisely the geometry of codon-anticodon base pair at the first two positions but monitor the third, or wobble position, less stringently, allowing diverse unusual base pairs to interact [58,59].

According to the energies of the codon-anticodon minihelices calculated following the Turner values [60], the Gly amino acid decoded by GGN codons belongs to the 'STRONG' group with high energy (average −4.6 kcal/mole) in their codon-anticodon duplexes [28]. Such high energy represents a risk of miscoding by binding to other GC-rich non-cognate aa-tRNAs. Therefore, the translation apparatus evolved to minimize the differential thermodynamic contributions of the codon-anticodon interactions. It results that the structures of tRNAs harboring too strong base-pairing binding capacity have been 'tuned down' during evolution for optimal and uniform translation. Consequently, the binding of natural tRNAs on the ribosome is very smilar whatever the GC/AU composition of the codon-anticodon duplex [27]. One critical feature that modulates the binding energy of codon-anticodon duplexes is the interaction between bases 32 and 38 in the tRNA anticodon loop. Classical Watson-Crick base pair constraints the anticodon loop structure and antagonize the canonical conformation of the codon-anticodon triplet, thereby weakening the corresponding binding energy. Watson-Crick base pairs in 32–38 are often associated with GC-rich anticodons, whereas weaker non-canonical pairs are

found in tRNAs with AU-rich anticodons where they allow optimal anticodon conformation for pairing with the codon. By this way, the differential thermodynamic contributions of all the codon-anticodon interactions is averaged, resulting in a rather uniform binding of tRNAs to the ribosome [28,61,62]. Here in our experiments the high binding energy of Gly codons inserted in the PK couldn't be balanced by a compensatory change in the IGR since it remained unchanged. This may explain that 2 G=C bp already reached an energy limit above which any additional Watson-Crick bp would decrease translation initiation efficiency. Above this limit, the PK probably possess too much energy compared to the external stabilizing grip energy for the ribosome required during translation, resulting in a possible rejection from the ribosomal decoding site.

#### 2.4.4. Frameshifting Activity of the IGR Programmed with GGU Codons

During translation, tRNAs move from the A site on the ribosome to the P site and then from the P site to the E site together with the mRNA. While moving, the codon-anticodon interactions remain intact through the A, P and E sites. A critical checkpoint for the maintenance of ribosome fidelity is located in the A-site where the geometry of codon–anticodon helix is tightly monitored by the three universally conserved and essential nts A1492, A1493, and G530 (bacterial numbering) of 16S rRNA [63,64]. This is achieved by intimate interactions with the minor groove of the first two bp of the codon-anticodon helix, thus sensing Watson-Crick base-pairing geometry and discriminating against near-cognate tRNA. The third (wobble) position of the codon is free to accommodate certain non-canonical base pairs. At this level, C1054 of 16S rRNA stacks against base 36 of the tRNA, which also interacts with G530, and ribosomal protein S12.

Additionally, natural modifications located in the anticodon loop (nt 37) and anticodon itself (nt 34), stabilize tRNA-codon interactions and increase decoding fidelity, or expand decoding capacity (nt 34). Downstream of the A-codon, nucleotides +7, +8 and +9 (the A of the AUG being nt+1) of the mRNA are held in place by a combination of hydrogen bonding and continuous aromatic base stacking involving, among others, C1054 from helix 34 of 16S rRNA, further stabilizing the network [65]. Altogether, data suggest that during elongation, the mRNA is stabilized and aligned at the entrance of the decoding center by a network of interactions with the ribosome [66,67].

Recently, a significant variation of this stabilizing network was observed in the cryo-EM structure of the IGR element of CrPV in complex with *K. lactis* ribosome [9]. Instead of the usual interaction between the C1054 (C1273 in *K. lactis*) and codon base 3, a unique pairing is observed with nt+1 of the first translated codon of the viral protein. This exceptional base pair results from the rotation of the base around its glycosidic bond causing the redirection of C1273 towards the first base of the first coding codon of the CrPV RNA [9]. Remarkably, with the exception of the IGR of PSIV [48], all the IGR sequences obtained to date exhibit a guanine residue at this position suggesting that the interaction C1273-G+1 may be mostly conserved in the dicistrovirus family. A sub-class of IGR elements synthesize in addition to the main ORF2 an overlapping hidden polypeptide (ORFx) in the +1 reading frame. Therefore it was proposed that the extra base pair C1273-G+1 might be an important regulatory element downstream of the ribosomal A site that could be involved in the frame selection by the IGR element [44,45,47]. In Israeli Acute Paralysis Virus (IAPV) IGR, the ORFx synthesis is initiated immediately downstream of the IGR, using a mechanism of frame selection probably dictated by the kinetics of tRNA binding in the first 0 or +1 frame codon as measured for the CrPV IGR [14].

Here in response to G+1 substitutions in the wild-type CrPV IGR, a decrease of translation of the 0 frame was observed (Figures 5 and 6). Translation of the +1 frame was also decreased but to a lesser extent, resulting in a global loss of fidelity in the frame maintenance as shown by the vertical stacked bar graphs (Figures 5B and 6B). Three constructs (AGG/CCU-C, GCC/GGU-A and GCC-GGU-U) exhibited frameshifting activity higher than 30% of the total R-Luc activity (Figures 4B and 5B).

Therefore, the C1273-G+1 interaction contributes to efficiency and fidelity of the translation process, assisting the PKI element in the presentation of the adjacent first codon in a proper conformation to the first incoming tRNA. The IGR IRES has a dynamic structure that is able to adopt upon ribosome-binding

all the conformations of the aminoacyl-tRNA from the initial binding to the fully accommodated stage after translocation into the P-site [12]. Our data and previous reports [9,12] clearly show that nt+1 plays a critical role when interacting with C1273 (C1054 in *E. coli* ribosomal RNA) in a conformation reminiscent of a hybrid tRNA state [44,45,47]. Although this interaction has never been observed in the large number of described ribosomal complexes with tRNAs or ASLs (for a full set of references [9]), one cannot exclude that C1273 (C1054) interacts with G+1 during standard decoding. C1273 may flip from the interaction with G+1 during the initial binding to stacking on the wobble base 34 as previously reported [50,51]. Another evidence of the importance of C1273 (C1054) during ribosome decoding is coming from early genetic studies. Mutants of C1054 with reduced translational accuracy were isolated, causing non-specific read-through of stop codons as well as enhanced +1 and −1 frameshifting [63].

Based on our systematic mutational analyses, the most likely scenario is that ribosomes recruited to the IGR initiate in the 0-frame or shift in the +1 frame depending on the codon-anticodon and downstream +1 nt. The presence of C1273 (C1054) would stabilize the conformation of PKI by interacting with a G in +1. Here evidence is presented which shows that the IGR activity is favored when it is programmed with a U-ending codon (AGG/CCU, GCC/GGU, ACC/GGU) and a G in +1. It includes the Gly GGU codon with G in +1. Our first finding that a G preferentially follows a GGU codon (Figure 3) can now be explained in terms of decoding efficiency and reading frame fidelity in the context of the PKI element of the IGR element. Indeed, although the IGR mimics the tRNA shape, it lacks some critical structural elements such as the base pair interaction between nts 32 and 38 which usually fine-tunes the loop structure and harmonize tRNA affinity for ribosomes [68–71] and the proofreading step on the ribosome by eEF1A. Additional experiments are now required to validate the conclusions using Gly codons located outside of the PKI in the coding sequence of R-Luc. This will allow testing natural modified tRNAs in the context of competition of a total tRNA extract, including the phenomenon of miscoding by near-cognate tRNAs when that was not the case with the IGR reprogrammed element.

## 3. Materials and Methods

### 3.1. Data Processing

Data processing was performed using a Python script. Coding DNA sequences (CDSs) from genomic sequences of 23 eukaryotic species were downloaded from the NCBI site ftp://ftp.ncbi.nlm.nih.gov/genomes/.

The first part of the script checked all CDSs of each species. It removed CDSs from the non-nuclear genome, annotated as pseudogenes and having a nucleotide length not a multiple of three. CDSs were translated taking into account the genetic code of each species and the translation exceptions like the selenocysteine UGA codon. CDSs were excluded if their protein sequence (i) did not begin with an AUG or an alternative initiator codon, (ii) did not end with a stop codon and (iii) contained a stop codon within their coding phase.From the DNA sequences of the retained CDS set, the second part of the script used a sliding window (window size of 6 bases and step size of 3 bases) to search four nucleotide patterns. These patterns correspond to the four glycine codons, GGU, GGG, GGC and GGA, followed by all base possibilities in position +1 and unknown bases in positions +2 and +3. For example, patterns GGTANN, GGTGNN, GGTCNN and GGTTNN were searched. The occurrence of the four given patterns was counted to calculate the relative percentage of each pattern. The output of the script was a csv file.

### 3.2. Construction of the IGR-luciferase Reporters

The PK mutants were constructed by PCR amplification of the CrPV IGR coding template using a sense primer (5'GGATATTAATACGACTCACTATAGG*CAAAAATGTGATCTTGCTT*3') appending the T7 RNA polymerase promoter (underlined sequence) to the IGR structure (italic) and an antisense primer (5'**CGAAGTATCTTGAAATGT***AGCAGGTAAATTTCTTAGGTTTTT CGACTACCTAATCTGAAA*3') where nts from the IGR are italicized, those forming the pseudo codon/anticodon are underlined and bold nts are complementary to Renilla Luciferase gene (Integrated DNA Technologies, Coralville, IA, USA). To generate the IGR variants, a PCR mixture containing 0.02 ng/µL of CrPV IGR-containing template plasmid, 0.2 µM of sense and antisense primer (Integrated DNA Technologies), 0.2 mM of each dNTP (ThermoFisher, Waltham, MA, USA), 0.04 U/µL of Phusion DNA polymerase (ThermoFisher) and the corresponding buffer at the recommended concentration were subjected to an initial denaturation step of 2 min at 95 °C followed by 25 cycles of: 30 s at 95 °C, 30 s at 55 °C, 1 min at 72 °C, and terminated by a final extension of 10 min at 72 °C. PCR products were fractionated on a 1% agarose Tris-acetate gel and the PCR product was excised from the gel and eluted by PCR Clean-Up System" kit (Promega, Madison, WI, USA) before quantification with a Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA). This PCR product was then used as primer in a second PCR reaction using as template the Renilla Luciferase gene (pRLuc). The second PCR was performed in the same conditions as above but using 0.02 ng/µL of pRLuc, 0.07 µM of the first PCR product (the region of the antisense primer complementary to R-Luc is in bold) and 0.2 µM of RevRLuc primer (5'GAATTATTGTTCATTTTTGAGAAC3'). To measure the luciferase activity resulting from a frameshift of +1 nucleotide, the second PCR amplification was performed on template pRLuc+1, which exhibits one extra A before residue Met 14 of Renilla Luciferase. Finally, the second PCR products were purified on a 1% agarose gel, and quantified as above. IGR-RLuc reporter mRNAs were synthesized from the second PCR product by in vitro transcription using recombinant T7 RNA polymerase as described previously [72,73]. After transcription, unincorporated nucleotides were trapped on a G-25 column. RNA transcripts were then phenol-extracted and ethanol precipitated. RNA pellets were resuspended in water and their concentration was determined by absorbance measurements. Figure A3 shows the RNA integrity of the constructs used in the study.

### 3.3. In vitro Translation in Rabbit Reticulocyte Lysates

Translation reactions were performed in RRL extracts as previously described [74]. Prior translation, the IGR RNA was heated at 95 °C for 3 min, then cooled to room temperature during 5 min and placed on ice. Translation reactions were incubated at 30 °C for 60 min and included 100 nM of each IGR-reporter mRNA transcript and 10.8 µCi [$^{35}$S]Met. Aliquots of translation reactions were analyzed by 15% SDS-PAGE (data not shown) and luciferase activities were measured with the Promega dual-luciferase assay, using a Promega luminometer. Figure A4 shows the radioactive translation products after separation by SDS PAGE and autoradiography.

## 4. Conclusions

Here we combined experimental and bioinformatics approaches to analyze the Glycine codon usage with the 3' downstream nucleotide (nt+1). A clear preference for a G in +1 was detected after the GGU codon. When a G residue was in +1 of an IGR construct reprogrammed with a GGU codon, the translational activity of the reporter R-Luc was improved in the 0 frame and frameshifting activity in the +1 frame dropped to a minimum. As GGU codons are read by tRNA$^{Gly}_{GCC}$, our data suggest that G34 together with G+1 is there to prevent +1 frameshifting. Altogether, the results show that translation efficiency of Gly GGU codon by tRNA$^{Gly}_{GCC}$ is modulated by the first nucleotide of the adjacent codon, acting simultaneously on the frameshifting activity.

Glycine is a member of a 4-codon box family whose two first nts are forming strong G=C pairings. Commonly, the Gly box is read by up to three different isoacceptors. What is remarkable in this box is that $tRNA^{Gly}_{A34CC}$ has been systematically banished from nearly all organisms, being replaced by $tRNA^{Gly}_{G34CC}$ with decoding properties for GGC and GGU codons. The other boxes usually exhibit a $tRNA_{ANN}$ that is subsequently deaminated in inosine with extended decoding properties for C, U and A-ending codons. The reason of the exceptional A34 sparing in the Gly box family is not clearly understood but several hypotheses have been proposed. For instance, it was shown that $tRNA^{Gly}_{A34CC}$ was an extremely poor substrate for the deaminase enzyme (adenosine deaminase acting on tRNA; ADAT) responsible for I34 formation [75]. This might be due to an unusual structure of the loop. Molecular dynamics (MD) simulations revealed that an A34 could be an important source of structural instability in the anticodon stem-loop of $tRNA^{Gly}$ [75]. Moreover, solution nuclear magnetic resonance (NMR) analyses on different $tRNA^{Gly}$ anticodon arms showed that they do not form the classical U-turn motif seen in tRNA anticodon loops [76]. Together, these results indicate that the singular $tRNA^{Gly}$ anticodon loop structure might be a potential cause for the selection of G34 versus I34 in $tRNA^{Gly}$. An unmodified A34 would only be able to decode U-ending codons while $tRNA^{Gly}_{G34CC}$ is able to decode G and U-ending codons. Beyond this idiosyncrasy, it is worth pointing out that evolution has assigned to $tRNA^{Gly}_{G34CC}$ and whole $tRNA^{Gly}$ family multiple extra-translational roles, including transcriptional regulation with the T-box regulatory elements [77] and cell wall biosynthesis [78].

**Author Contributions:** G.E. and F.M. conceived and supervised the study; A.J., A.T. and L.S. conducted the experiments; Software, L.D.; Data analysis, F.M. and G.E.; G.E. wrote the manuscript; Funding acquisition, G.E. and F.M. All authors read and approved the manuscript.

**Funding:** This work was supported by 'Centre National de la Recherche Scientifique' (CNRS) and funded by 'Agence Nationale pour la Recherche' (ANR) [H4translation, ANR-17-CE11-0024-01] and [Ribofluidix, ANR-17-CE12-0025-01].

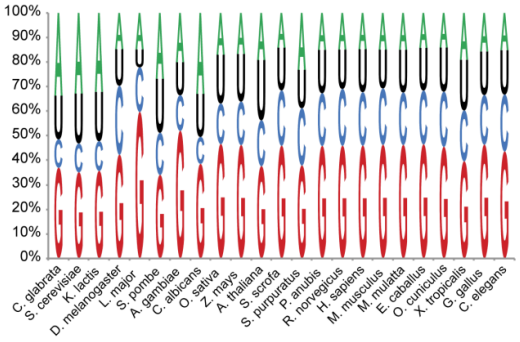
**Acknowledgments:** The authors wish to thank Eric Westhof for precious discussions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

| | |
|---|---|
| ASL | Anticodon Stem Loop |
| CDS | Coding DNA sequence |
| CrPV | Cricket Paralysis Virus |
| IGR | InterGenic Region |
| IRES | Internal Ribosomal Entry Site |
| nt | nucleotide |
| ORF | Open Reading Frame |
| PKI | Pseudoknot I |
| R-Luc | Renilla Luciferase |
| RRL | Rabbit Reticulocyte Lysate |

## Appendix A




| Organism name | Codon number Motif searched | Motif searched Percentage | Motif searched Percentage | Codon number | Percentage Codon number | Codon number Motif searched | Motif searched Percentage | Percentage Codon number |
|---|---|---|---|---|---|---|---|---|
| Strongylocentrotus_purpuratus | GGAANN 110112 | 27.017 | GGAGNN | 154588 | 37.929 | GGACNN 73233 | 17.968 | GGATNN 69636 17.086 |
| Leishmania_major | GGAANN 8441 | 24.750 | GGAGNN | 13118 | 38.464 | GGACNN 8085 | 23.706 | GGATNN 4461 13.080 |
| Saccharomyces_cerevisiae | GGAANN 12458 | 37.990 | GGAGNN | 7656 | 23.346 | GGACNN 4731 | 14.427 | GGATNN 7948 24.237 |
| Candida_glabrata | GGAANN 9092 | 34.540 | GGAGNN | 6435 | 24.446 | GGACNN 4153 | 15.777 | GGATNN 6643 25.236 |
| Kluyveromyces_lactis | GGAANN 10361 | 33.386 | GGAGNN | 7217 | 23.255 | GGACNN 4518 | 14.558 | GGATNN 8938 28.801 |
| Candida_albicans | GGAANN 12354 | 28.547 | GGAGNN | 11072 | 25.585 | GGACNN 5309 | 12.268 | GGATNN 14541 33.601 |
| Schizosaccharomyces_pombe | GGAANN 14781 | 38.998 | GGAGNN | 8839 | 23.321 | GGACNN 5463 | 14.413 | GGATNN 8819 23.268 |
| Anopheles_gambiae | GGAANN 16384 | 24.731 | GGAGNN | 19009 | 28.693 | GGACNN 17589 | 26.549 | GGATNN 13268 20.027 |
| Drosophila_melanogaster | GGAANN 88124 | 24.286 | GGAGNN | 118498 | 32.656 | GGACNN 83517 | 23.016 | GGATNN 72726 20.042 |
| Homo_sapiens | GGAANN 403928 | 30.233 | GGAGNN | 519171 | 38.859 | GGACNN 246856 | 18.477 | GGATNN 166074 12.430 |
| Mus_musculus | GGAANN 249647 | 27.964 | GGAGNN | 357757 | 40.074 | GGACNN 179996 | 20.162 | GGATNN 105352 11.801 |
| Oryctolagus_cuniculus | GGAANN 107397 | 29.859 | GGAGNN | 141285 | 39.281 | GGACNN 66917 | 18.605 | GGATNN 44079 12.255 |
| Macaca_mulatta | GGAANN 192523 | 30.459 | GGAGNN | 244788 | 38.728 | GGACNN 116884 | 18.492 | GGATNN 77871 12.320 |
| Papio_anubis | GGAANN 234407 | 30.506 | GGAGNN | 298408 | 38.836 | GGACNN 141211 | 18.378 | GGATNN 94361 12.280 |
| Rattus_norvegicus | GGAANN 161721 | 27.709 | GGAGNN | 235675 | 40.381 | GGACNN 118505 | 20.305 | GGATNN 67734 11.606 |
| Sus_scrofa | GGAANN 211041 | 29.917 | GGAGNN | 275803 | 39.098 | GGACNN 131903 | 18.699 | GGATNN 86664 12.286 |
| Equus_caballus | GGAANN 221185 | 30.083 | GGAGNN | 290024 | 39.445 | GGACNN 136175 | 18.521 | GGATNN 87873 11.951 |
| Caenorhabditis_elegans | GGAANN 148247 | 34.889 | GGAGNN | 127854 | 30.090 | GGACNN 63801 | 15.015 | GGATNN 85009 20.006 |
| Arabidopsis_thaliana | GGAANN 162643 | 33.487 | GGAGNN | 168824 | 34.759 | GGACNN 66129 | 13.615 | GGATNN 88097 18.138 |
| Oryza_sativa | GGAANN 92343 | 31.322 | GGAGNN | 101342 | 34.375 | GGACNN 42279 | 14.341 | GGATNN 58850 19.962 |
| Zea_mays | GGAANN 116930 | 30.905 | GGAGNN | 130409 | 34.468 | GGACNN 60367 | 15.955 | GGATNN 70645 18.672 |
| Gallus_gallus | GGAANN 179007 | 29.860 | GGAGNN | 233128 | 38.887 | GGACNN 107260 | 17.892 | GGATNN 80099 13.361 |
| Xenopus_tropicalis | GGAANN 166459 | 32.656 | GGAGNN | 177528 | 34.827 | GGACNN 86203 | 16.911 | GGATNN 79547 15.605 |
| Strongylocentrotus_purpuratus | GGCANN 101229 | 40.419 | GGCGNN | 51855 | 20.705 | GGCCNN 43303 | 17.290 | GGCTNN 54060 21.585 |
| Leishmania_major | GGCANN 52412 | 29.796 | GGCGNN | 70801 | 40.250 | GGCCNN 26745 | 15.204 | GGCTNN 25946 14.750 |
| Saccharomyces_cerevisiae | GGCANN 9467 | 33.001 | GGCGNN | 7572 | 26.395 | GGCCNN 5045 | 17.586 | GGCTNN 6603 23.017 |
| Candida_glabrata | GGCANN 8206 | 33.502 | GGCGNN | 4930 | 20.127 | GGCCNN 4695 | 19.168 | GGCTNN 6663 27.203 |
| Kluyveromyces_lactis | GGCANN 5563 | 33.425 | GGCGNN | 4048 | 24.323 | GGCCNN 3123 | 18.765 | GGCTNN 3909 23.487 |
| Candida_albicans | GGCANN 5467 | 41.857 | GGCGNN | 3004 | 23.000 | GGCCNN 1798 | 13.766 | GGCTNN 2792 21.377 |
| Schizosaccharomyces_pombe | GGCANN 6062 | 31.212 | GGCGNN | 4995 | 25.718 | GGCCNN 3422 | 17.619 | GGCTNN 4943 25.451 |
| Anopheles_gambiae | GGCANN 59216 | 35.697 | GGCGNN | 57836 | 34.865 | GGCCNN 25669 | 15.474 | GGCTNN 23166 13.965 |
| Drosophila_melanogaster | GGCANN 195313 | 36.713 | GGCGNN | 175662 | 33.019 | GGCCNN 71767 | 13.490 | GGCTNN 89257 16.778 |
| Homo_sapiens | GGCANN 495190 | 31.913 | GGCGNN | 213924 | 13.787 | GGCCNN 459229 | 29.596 | GGCTNN 383337 24.705 |
| Mus_musculus | GGCANN 349935 | 33.129 | GGCGNN | 130925 | 12.395 | GGCCNN 300489 | 28.448 | GGCTNN 274931 26.028 |
| Oryctolagus_cuniculus | GGCANN 142182 | 30.588 | GGCGNN | 79122 | 17.022 | GGCCNN 133280 | 28.673 | GGCTNN 110251 23.718 |
| Macaca_mulatta | GGCANN 236155 | 32.391 | GGCGNN | 103961 | 14.259 | GGCCNN 212052 | 29.085 | GGCTNN 176898 24.264 |
| Papio_anubis | GGCANN 284540 | 32.297 | GGCGNN | 123990 | 14.074 | GGCCNN 258838 | 29.380 | GGCTNN 213636 24.249 |
| Rattus_norvegicus | GGCANN 229917 | 32.892 | GGCGNN | 88941 | 12.724 | GGCCNN 199504 | 28.541 | GGCTNN 180640 25.843 |
| Sus_scrofa | GGCANN 297403 | 30.601 | GGCGNN | 164425 | 16.919 | GGCCNN 285005 | 29.326 | GGCTNN 225029 23.154 |
| Equus_caballus | GGCANN 296314 | 31.055 | GGCGNN | 154691 | 16.212 | GGCCNN 279002 | 29.241 | GGCTNN 224150 23.492 |
| Caenorhabditis_elegans | GGCANN 23651 | 26.586 | GGCGNN | 28790 | 32.363 | GGCCNN 16749 | 18.828 | GGCTNN 19769 22.223 |
| Arabidopsis_thaliana | GGCANN 58312 | 31.911 | GGCGNN | 45196 | 24.733 | GGCCNN 31824 | 17.416 | GGCTNN 47401 25.940 |
| Oryza_sativa | GGCANN 127613 | 25.924 | GGCGNN | 198764 | 40.378 | GGCCNN 91859 | 18.661 | GGCTNN 74021 15.037 |
| Zea_mays | GGCANN 172607 | 27.420 | GGCGNN | 231547 | 36.783 | GGCCNN 126934 | 20.165 | GGCTNN 98400 15.632 |
| Gallus_gallus | GGCANN 209008 | 36.755 | GGCGNN | 81211 | 14.281 | GGCCNN 134414 | 23.637 | GGCTNN 144021 25.327 |
| Xenopus_tropicalis | GGCANN 128532 | 40.360 | GGCGNN | 35207 | 11.055 | GGCCNN 72473 | 22.757 | GGCTNN 82250 25.827 |
| Strongylocentrotus_purpuratus | GGGANN 70499 | 34.293 | GGGGNN | 62791 | 30.543 | GGGCNN 34729 | 16.893 | GGGTNN 37560 18.270 |
| Leishmania_major | GGGANN 11838 | 18.924 | GGGGNN | 18850 | 30.133 | GGGCNN 20802 | 33.254 | GGGTNN 11065 17.688 |
| Saccharomyces_cerevisiae | GGGANN 5636 | 31.704 | GGGGNN | 3722 | 20.937 | GGGCNN 3488 | 19.621 | GGGTNN 4931 27.738 |
| Candida_glabrata | GGGANN 4738 | 32.379 | GGGGNN | 2735 | 18.691 | GGGCNN 2826 | 19.313 | GGGTNN 4334 29.618 |
| Kluyveromyces_lactis | GGGANN 5522 | 35.697 | GGGGNN | 2729 | 17.642 | GGGCNN 2141 | 13.841 | GGGTNN 5077 32.820 |
| Candida_albicans | GGGANN 8081 | 36.264 | GGGGNN | 4744 | 21.289 | GGGCNN 1477 | 6.628 | GGGTNN 7982 35.819 |
| Schizosaccharomyces_pombe | GGGANN 3431 | 32.068 | GGGGNN | 2348 | 21.946 | GGGCNN 2158 | 20.170 | GGGTNN 2762 25.815 |
| Anopheles_gambiae | GGGANN 10600 | 17.199 | GGGGNN | 14893 | 24.165 | GGGCNN 23564 | 38.234 | GGGTNN 12574 20.402 |
| Drosophila_melanogaster | GGGANN 21597 | 23.445 | GGGGNN | 32925 | 35.742 | GGGCNN 26199 | 28.441 | GGGTNN 11397 12.372 |
| Homo_sapiens | GGGANN 354536 | 29.565 | GGGGNN | 414679 | 34.580 | GGGCNN 311271 | 25.957 | GGGTNN 118695 9.898 |
| Mus_musculus | GGGANN 248075 | 31.239 | GGGGNN | 259560 | 32.685 | GGGCNN 198821 | 25.037 | GGGTNN 87666 11.039 |
| Oryctolagus_cuniculus | GGGANN 98708 | 29.224 | GGGGNN | 115032 | 34.057 | GGGCNN 90706 | 26.855 | GGGTNN 33319 9.865 |
| Macaca_mulatta | GGGANN 162407 | 29.438 | GGGGNN | 191731 | 34.753 | GGGCNN 142850 | 25.893 | GGGTNN 54708 9.916 |
| Papio_anubis | GGGANN 201394 | 29.733 | GGGGNN | 231843 | 34.228 | GGGCNN 176250 | 26.021 | GGGTNN 67862 10.019 |
| Rattus_norvegicus | GGGANN 166729 | 31.473 | GGGGNN | 172600 | 32.582 | GGGCNN 131582 | 24.839 | GGGTNN 58836 11.106 |
| Sus_scrofa | GGGANN 200261 | 27.839 | GGGGNN | 257254 | 35.762 | GGGCNN 192832 | 26.806 | GGGTNN 69003 9.592 |
| Equus_caballus | GGGANN 206227 | 28.839 | GGGGNN | 251646 | 35.190 | GGGCNN 188816 | 26.404 | GGGTNN 68415 9.567 |
| Caenorhabditis_elegans | GGGANN 18486 | 32.750 | GGGGNN | 10831 | 19.189 | GGGCNN 16244 | 28.778 | GGGTNN 10884 19.282 |
| Arabidopsis_thaliana | GGGANN 94363 | 44.677 | GGGGNN | 40291 | 19.076 | GGGCNN 30852 | 14.607 | GGGTNN 45707 21.640 |
| Oryza_sativa | GGGANN 95470 | 32.658 | GGGGNN | 73793 | 25.243 | GGGCNN 70854 | 24.237 | GGGTNN 52219 17.863 |
| Zea_mays | GGGANN 114790 | 30.036 | GGGGNN | 99769 | 26.105 | GGGCNN 100245 | 26.230 | GGGTNN 67376 17.629 |
| Gallus_gallus | GGGANN 151351 | 31.925 | GGGGNN | 148282 | 31.277 | GGGCNN 125249 | 26.419 | GGGTNN 49207 10.379 |
| Xenopus_tropicalis | GGGANN 105341 | 32.240 | GGGGNN | 109862 | 33.623 | GGGCNN 71265 | 21.811 | GGGTNN 40276 12.326 |
| Strongylocentrotus_purpuratus | GGTANN 81951 | 23.177 | GGTGNN | 133884 | 37.864 | GGTCNN 82581 | 23.355 | GGTTNN 55177 15.605 |
| Leishmania_major | GGTANN 9772 | 14.845 | GGTGNN | 39278 | 59.668 | GGTCNN 11999 | 18.228 | GGTTNN 4779 7.260 |
| Saccharomyces_cerevisiae | GGTANN 21697 | 32.850 | GGTGNN | 23182 | 35.098 | GGTCNN 7739 | 11.717 | GGTTNN 13431 20.335 |
| Candida_glabrata | GGTANN 24575 | 33.694 | GGTGNN | 26955 | 36.957 | GGTCNN 8435 | 11.565 | GGTTNN 12971 17.784 |
| Kluyveromyces_lactis | GGTANN 20062 | 32.142 | GGTGNN | 22260 | 35.663 | GGTCNN 7547 | 12.091 | GGTTNN 12548 20.103 |
| Candida_albicans | GGTANN 23701 | 32.986 | GGTGNN | 27668 | 38.507 | GGTCNN 7890 | 10.981 | GGTTNN 12592 17.525 |
| Schizosaccharomyces_pombe | GGTANN 13102 | 26.810 | GGTGNN | 16618 | 34.005 | GGTCNN 8408 | 17.205 | GGTTNN 10742 21.981 |
| Anopheles_gambiae | GGTANN 20113 | 20.071 | GGTGNN | 52089 | 51.980 | GGTCNN 14542 | 14.512 | GGTTNN 13465 13.437 |
| Drosophila_melanogaster | GGTANN 40390 | 14.775 | GGTGNN | 115468 | 42.238 | GGTCNN 76901 | 28.130 | GGTTNN 40614 14.857 |
| Homo_sapiens | GGTANN 124557 | 14.788 | GGTGNN | 387171 | 45.965 | GGTCNN 183284 | 21.760 | GGTTNN 147296 17.487 |
| Mus_musculus | GGTANN 88335 | 14.958 | GGTGNN | 273146 | 46.251 | GGTCNN 135929 | 23.017 | GGTTNN 93157 15.774 |
| Oryctolagus_cuniculus | GGTANN 30267 | 14.105 | GGTGNN | 97541 | 45.457 | GGTCNN 48227 | 22.475 | GGTTNN 38543 17.962 |
| Macaca_mulatta | GGTANN 60629 | 14.973 | GGTGNN | 187217 | 46.236 | GGTCNN 86596 | 21.386 | GGTTNN 70472 17.404 |
| Papio_anubis | GGTANN 73018 | 15.028 | GGTGNN | 222614 | 45.816 | GGTCNN 104533 | 21.514 | GGTTNN 85726 17.643 |
| Rattus_norvegicus | GGTANN 58527 | 14.884 | GGTGNN | 180512 | 45.905 | GGTCNN 91175 | 23.186 | GGTTNN 63017 16.025 |
| Sus_scrofa | GGTANN 62925 | 14.142 | GGTGNN | 204335 | 45.923 | GGTCNN 100383 | 22.560 | GGTTNN 77311 17.375 |
| Equus_caballus | GGTANN 65693 | 14.236 | GGTGNN | 214228 | 46.426 | GGTCNN 101507 | 21.998 | GGTTNN 80014 17.340 |
| Caenorhabditis_elegans | GGTANN 22328 | 15.153 | GGTGNN | 64214 | 43.579 | GGTCNN 34206 | 23.214 | GGTTNN 26602 18.054 |
| Arabidopsis_thaliana | GGTANN 84668 | 19.169 | GGTGNN | 166014 | 37.586 | GGTCNN 82527 | 18.684 | GGTTNN 108486 24.561 |
| Oryza_sativa | GGTANN 50292 | 17.034 | GGTGNN | 137601 | 46.605 | GGTCNN 48667 | 16.483 | GGTTNN 58688 19.878 |
| Zea_mays | GGTANN 65234 | 16.695 | GGTGNN | 180484 | 46.190 | GGTCNN 69149 | 17.697 | GGTTNN 75872 19.418 |
| Gallus_gallus | GGTANN 60771 | 15.700 | GGTGNN | 179453 | 46.361 | GGTCNN 78261 | 20.219 | GGTTNN 68591 17.720 |
| Xenopus_tropicalis | GGTANN 56775 | 19.068 | GGTGNN | 117810 | 39.566 | GGTCNN 61798 | 20.755 | GGTTNN 61369 20.611 |
| Sums | 9943341 | | 11302477 | | | 7975808 | | 6024626 |

**Figure A1.** Distribution of nucleotide +1 downstream of GGU Gly codons in each of the 23 eukaryotic genomes. 35,246,252 Gly codons were analysed. Results are expressed as percentages. The *x*-axis corresponds

to the organisms while the *y*-axis refers to the occurrence of each nucleotide at position +1. Details of the analysis are shown below the graph.

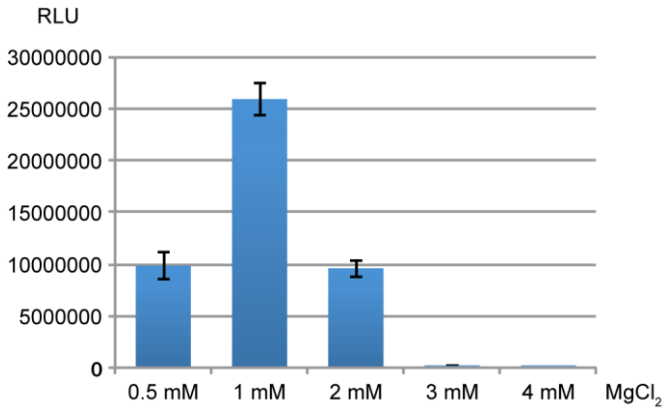



**Figure A2.** The effect on translation of the $Mg^{2+}$ concentration added to the extracts. R-Luc activity was measured in the 0 frame. Translation efficiencies are represented as raw bioluminescence activity (Relative Light Units or RLU) and resulted from at least 3 independent experiments. Standard deviations are shown.

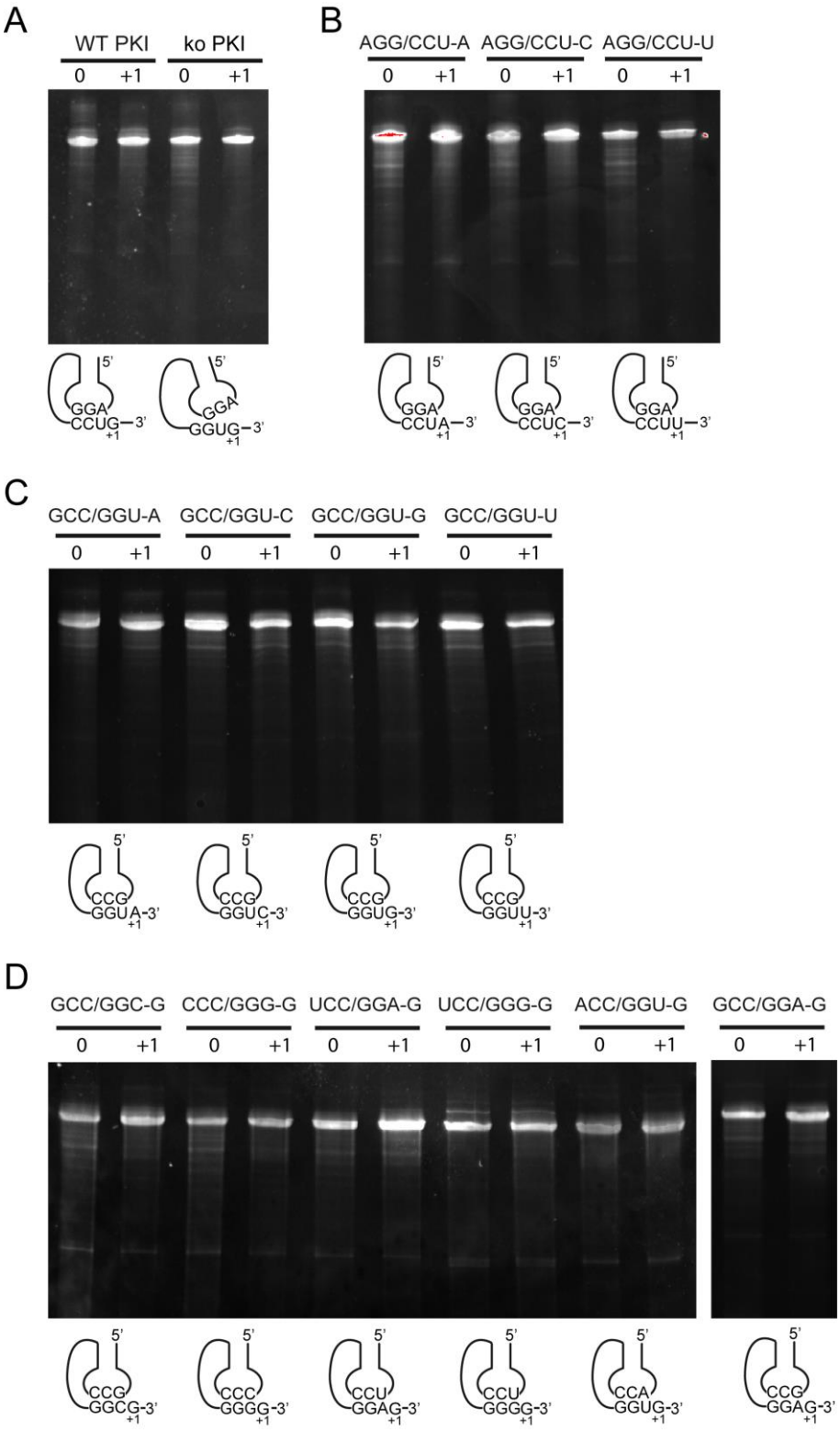



**Figure A3.** Control of the integrity of the RNA constructs used in the study. RNA samples (0.4–0.6 μg) were fractionated by 4% denaturing PAGE and stained with ethidium bromide. (**A**) RNAs used in Figure 4; (**B**) RNAs used in Figure 5; (**C**) RNAs used in Figure 6; (**D**) RNAs used in Figure 7.

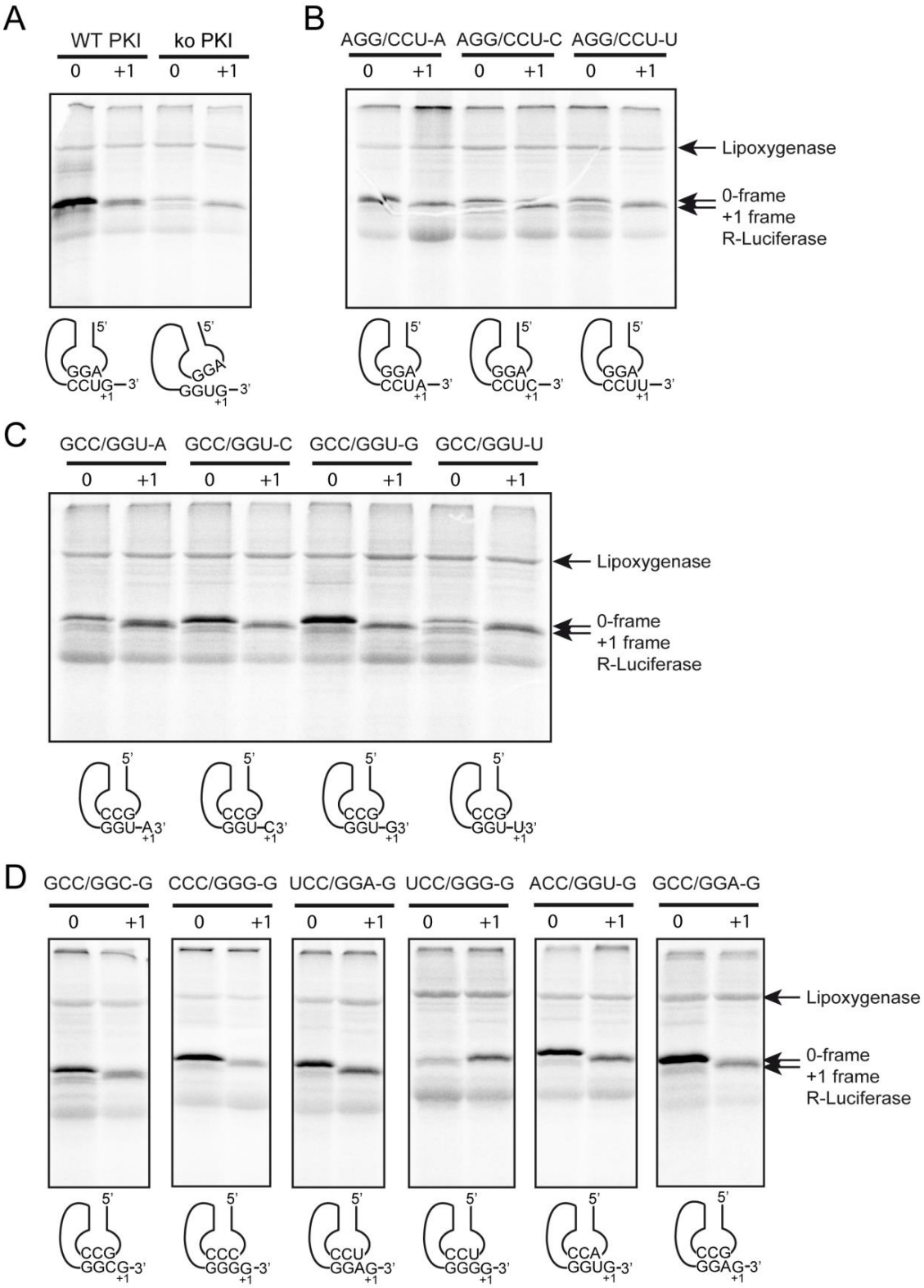



**Figure A4.** Analysis of the IGR-dependent translation products in vitro. For each construct, the 0- and +1-frame products were analysed. The reporter constructs were incubated in RRL translation extracts at 30 °C for 60 min in the presence of [$^{35}$S]Met. Translation experiments were performed in untreated RRL containing endogenous lipoxygenase mRNA as well as globin mRNA in large amount (the globin band is not shown). Translation products were visualized by autoradiography after resolving on a 10% SDS PAGE.

## References


1. Pape, T.; Wintermeyer, W.; Rodnina, M. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J.* **1999**, *18*, 3800–3807. [CrossRef] [PubMed]
2. Rodnina, M.V.; Wintermeyer, W. Fidelity of Aminoacyl-tRNA Selection on the Ribosome: Kinetic and Structural Mechanisms. *Annu. Rev. Biochem.* **2001**, *70*, 415–435. [CrossRef] [PubMed]
3. Rodnina, M.V.; Wintermeyer, W. Ribosome fidelity: tRNA discrimination, proofreading and induced fit. *Trends Biochem. Sci.* **2001**, *26*, 124–130. [CrossRef]
4. Sasaki, J.; Nakashima, N. Translation initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus in vitro. *J. Virol.* **1999**, *73*, 1219–1226. [PubMed]
5. Wilson, J.E.; Pestova, T.V.; Hellen, C.U.; Sarnow, P. Initiation of protein synthesis from the A site of the ribosome. *Cell* **2000**, *102*, 511–520. [CrossRef]
6. Fernández, I.S.; Bai, X.C.; Murshudov, G.; Scheres, S.H.W.; Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **2014**, *157*, 823–831. [CrossRef] [PubMed]
7. Costantino, D.A.; Pfingsten, J.S.; Rambo, R.P.; Kieft, J.S. tRNA–mRNA mimicry drives translation initiation from a viral IRES. *Nat. Struct. Mol. Biol.* **2008**, *15*, 57–64. [CrossRef]
8. Koh, C.S.; Brilot, A.F.; Grigorieff, N.; Korostelev, A.A. Taura syndrome virus IRES initiates translation by binding its tRNA-mRNA-like structural element in the ribosomal decoding center. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 9139–9144. [CrossRef]
9. Murray, J.; Savva, C.G.; Shin, B.S.; Dever, T.E.; Ramakrishnan, V.; Fernández, I.S. Structural characterization of ribosome recruitment and translocation by type IV IRES. *Elife* **2016**, *5*, e13567. [CrossRef]
10. Hussain, T.; Llácer, J.L.; Fernández, I.S.; Munoz, A.; Martin-Marcos, P.; Savva, C.G.; Lorsch, J.R.; Hinnebusch, A.G.; Ramakrishnan, V. Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* **2014**, *159*, 597–607. [CrossRef]
11. Erzberger, J.P.; Stengel, F.; Pellarin, R.; Zhang, S.; Schaefer, T.; Aylett, C.H.S.; Cimermančič, P.; Boehringer, D.; Sali, A.; Aebersold, R.; et al. Molecular architecture of the 40S·eIF1·eIF3 translation initiation complex. *Cell* **2014**, *158*, 1123–1135. [CrossRef] [PubMed]
12. Muhs, M.; Hilal, T.; Mielke, T.; Skabkin, M.A.; Sanbonmatsu, K.Y.; Pestova, T.V.; Spahn, C.M.T. Cryo-EM of ribosomal 80s complexes with termination factors reveals the translocated cricket paralysis virus IRES. *Mol. Cell* **2015**, *57*, 422–433. [CrossRef] [PubMed]
13. Pisareva, V.P.; Pisarev, A.V.; Fernández, I.S. Dual tRNA mimicry in the Cricket Paralysis Virus IRES uncovers an unexpected similarity with the Hepatitis C Virus IRES. *Elife* **2018**, *7*, e34062. [CrossRef] [PubMed]
14. Petrov, A.; Grosely, R.; Chen, J.; O'Leary, S.E.; Puglisi, J.D. Multiple Parallel Pathways of Translation Initiation on the CrPV IRES. *Mol. Cell* **2016**, *62*, 92–103. [CrossRef] [PubMed]
15. Kurland, C.G.; Ehrenberg, M. Growth-optimizing accuracy of gene expression. *Annu. Rev. Biophys. Biophys. Chem.* **1987**, *16*, 291–317. [CrossRef]
16. Farabaugh, P.J. Programmed translational frameshifting. *Microbiol. Rev.* **1996**, *60*, 103–134. [CrossRef]
17. Namy, O.; Moran, S.J.; Stuart, D.I.; Gilbert, R.J.; Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **2006**, *441*, 244–247. [CrossRef]
18. Atkins, J.F.; Loughran, G.; Bhatt, P.R.; Firth, A.E.; Baranov, P.V. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **2016**, *44*, 7007–7078. [CrossRef]
19. Atkins, J.F.; Gesteland, R.F.; Reid, B.R.; Anderson, C.W. Normal tRNAs promote ribosomal frameshifting. *Cell* **1979**, *18*, 1119–1131. [CrossRef]
20. Björk, G.R.; Wikström, P.M.; Byström, A.S. Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. *Science* **1989**, *244*, 986–989. [CrossRef]
21. Hagervall, T.G.; Tuohy, T.M.; Atkins, J.F.; Björk, G.R. Deficiency of 1-methylguanosine in tRNA from Salmonella typhimurium induces frameshifting by quadruplet translocation. *J. Mol. Biol.* **1993**, *232*, 756–765. [CrossRef] [PubMed]
22. Qian, Q.; Björk, G.R. Structural alterations far from the anticodon of the tRNAProGGG of Salmonella typhimurium induce +1 frameshifting at the peptidyl-site. *J. Mol. Biol.* **1997**, *273*, 978–992. [CrossRef] [PubMed]

23. Hong, S.; Sunita, S.; Maehigashi, T.; Hoffer, E.D.; Dunkle, J.A.; Dunham, C.M. Mechanism of tRNA-mediated +1 ribosomal frameshifting. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 11226–11231. [CrossRef] [PubMed]
24. Maehigashi, T.; Dunkle, J.A.; Miles, S.J.; Dunham, C.M. Structural insights into +1 frameshifting promoted by expanded or modification-deficient anticodon stem loops. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 12740–12745. [CrossRef] [PubMed]
25. Ledoux, S.; Olejniczak, M.; Uhlenbeck, O.C. A sequence element that tunes Escherichia coli tRNA(Ala)(GGC) to ensure accurate decoding. *Nat. Struct. Mol. Biol.* **2009**, *16*, 359–364. [CrossRef] [PubMed]
26. Murakami, H.; Ohta, A.; Suga, H. Bases in the anticodon loop of tRNAAlaGGC prevent misreading. *Nat. Struct. Mol. Biol.* **2009**, *16*, 353–358. [CrossRef] [PubMed]
27. Olejniczak, M.; Uhlenbeck, O.C. tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie* **2006**, *88*, 943–950. [CrossRef]
28. Grosjean, H.; Westhof, E. An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res.* **2016**, *44*, 8020–8040. [CrossRef]
29. Chan, P.P.; Lowe, T.M. GtRNAdb 2.0: An expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **2016**, *44*, D184–D189. [CrossRef]
30. Rogalski, M.; Karcher, D.; Bock, R. Superwobbling facilitates translation with reduced tRNA sets. *Nat. Struct. Mol. Biol.* **2008**, *15*, 192–198. [CrossRef]
31. Curran, J.F. Decoding with the A:I wobble pair is inefficient. *Nucleic Acids Res.* **1995**, *23*, 683–688. [CrossRef]
32. Pak, D.; Kim, Y.; Burton, Z.F. Aminoacyl-tRNA synthetase evolution and sectoring of the genetic code. *Transcription* **2018**, *9*, 1–15. [CrossRef]
33. Gaber, R.F.; Mathison, L.; Edelman, I.; Culbertson, M.R. Frameshift Suppression in *Saccharomyces cerevisiae* VI. Complete Genetic Map of Twenty-Five Suppressor Genes. *Genetics* **1983**, *103*, 389–407.
34. Mathew, S.F.; Crowe-McAuliffe, C.; Graves, R.; Cardno, T.S.; McKinney, C.; Poole, E.S.; Tate, W.P. The highly conserved codon following the slippery sequence supports—1 frameshift efficiency at the HIV-1 frameshift site. *PLoS ONE* **2015**, *10*, e0122176. [CrossRef]
35. Shehu-Xhilaga, M.; Crowe, S.M.; Mak, J. Maintenance of the Gag/Gag-Pol Ratio Is Important for Human Immunodeficiency Virus Type 1 RNA Dimerization and Viral Infectivity. *J. Virol.* **2001**, *75*, 1834–1841. [CrossRef]
36. Belew, A.T.; Meskauskas, A.; Musalgaonkar, S.; Advani, V.M.; Sulima, S.O.; Kasprzak, W.K.; Shapiro, B.A.; Dinman, J.D. Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway. *Nature* **2014**, *512*, 265–269. [CrossRef]
37. Advani, V.M.; Belew, A.T.; Dinman, J.D. Yeast telomere maintenance is globally controlled by programmed ribosomal frameshifting and the nonsense-mediated mRNA decay pathway. *Translation* **2013**, *1*, e24418. [CrossRef]
38. Sharma, V.; Prere, M.F.; Canal, I.; Firth, A.E.; Atkins, J.F.; Baranov, P.V.; Fayet, O. Analysis of tetra- and hepta-nucleotides motifs promoting -1 ribosomal frameshifting in Escherichia coli. *Nucleic Acids Res.* **2014**, *42*, 7210–7225. [CrossRef]
39. Nakamura, Y.; Gojobori, T.; Ikemura, T. Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids Res.* **2000**, *28*, 292. [CrossRef]
40. Murphy, F.V.; Ramakrishnan, V. Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1251–1252. [CrossRef]
41. Grosjean, H.J.; de Henau, S.; Crothers, D.M. On the physical basis for ambiguity in genetic coding interactions. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 610–614. [CrossRef]
42. Grosjean, H.; de Crécy-Lagard, V.; Marck, C. Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes. *FEBS Lett.* **2010**, *584*, 252–264. [CrossRef]
43. Komine, Y.; Kitabatake, M.; Yokogawa, T.; Nishikawa, K.; Inokuchi, H. A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 9223–9227. [CrossRef]
44. Kerr, C.H.; Wang, Q.S.; Moon, K.M.; Keatings, K.; Allan, D.W.; Foster, L.J.; Jan, E. IRES-dependent ribosome repositioning directs translation of a +1 overlapping ORF that enhances viral infection. *Nucleic Acids Res.* **2018**, *46*, 11952–11967. [CrossRef]
45. Ren, Q.; Au, H.H.T.; Wang, Q.S.; Lee, S.; Jan, E. Structural determinants of an internal ribosome entry site that direct translational reading frame selection. *Nucleic Acids Res.* **2014**, *42*, 9366–9382. [CrossRef]

46. Au, H.H.; Cornilescu, G.; Mouzakis, K.D.; Ren, Q.; Burke, J.E.; Lee, S.; Butcher, S.E.; Jan, E. Global shape mimicry of tRNA within a viral internal ribosome entry site mediates translational reading frame selection. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6446–E6455. [CrossRef]
47. Ren, Q.; Wang, Q.S.; Firth, A.E.; Chan, M.M.Y.; Gouw, J.W.; Guarna, M.M.; Foster, L.J.; Atkins, J.F.; Jan, E. Alternative reading frame selection mediated by a tRNA-like domain of an internal ribosome entry site. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E630–E639. [CrossRef]
48. Kieft, J.S. Comparing the three-dimensional structures of Dicistroviridae IGR IRES RNAs with other viral RNA structures. *Virus Res.* **2009**, *139*, 148–156. [CrossRef]
49. Varshavsky, A. The N-end rule. *Cell* **1992**, *69*, 725–735. [CrossRef]
50. Westhof, E. Isostericity and tautomerism of base pairs in nucleic acids. *FEBS Lett.* **2014**, *588*, 2464–2469. [CrossRef]
51. Westhof, E.; Yusupov, M.; Yusupova, G. Recognition of Watson-Crick base pairs: Constraints and limits due to geometric selection and tautomerism. *F1000Prime Rep.* **2014**, *6*, 19. [CrossRef]
52. Murphy, F.V.; Ramakrishnan, V.; Malkiewicz, A.; Agris, P.F. The role of modifications in codon discrimination by tRNALysUUU. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186–1191. [CrossRef]
53. Vendeix, F.A.P.; Murphy, F.V.; Cantara, W.A.; Leszczyńska, G.; Gustilo, E.M.; Sproat, B.; Malkiewicz, A.; Agris, P.F. Human tRNA(Lys3)(UUU) is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism. *J. Mol. Biol.* **2012**, *416*, 467–485. [CrossRef]
54. Kurata, S.; Weixlbaumer, A.; Ohtsuki, T.; Shimazaki, T.; Wada, T.; Kirino, Y.; Takai, K.; Watanabe, K.; Ramakrishnan, V.; Suzuki, T. Modified Uridines with C5-methylene Substituents at the First Position of the tRNA Anticodon Stabilize U·G Wobble Pairing during Decoding. *J. Biol. Chem.* **2008**, *283*, 18801–18811. [CrossRef]
55. Weixlbaumer, A.; Murphy, F.V.; Dziergowska, A.; Malkiewicz, A.; Vendeix, F.A.P.; Agris, P.F.; Ramakrishnan, V. Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nat. Struct. Mol. Biol.* **2007**, *14*, 498–502. [CrossRef]
56. Ito, Y.; Sone, Y.; Mizutani, T. Stability of non-Watson-Crick G-A/A-G base pair in synthetic DNA and RNA oligonucleotides. *Mol. Biol. Rep.* **2004**, *31*, 31–36. [CrossRef]
57. Olson, W.K.; Li, S.; Kaukonen, T.; Colasanti, A.V.; Xin, Y.; Lu, X.J. Effects of Noncanonical Base Pairing on RNA Folding: Structural Context and Spatial Arrangements of G·A Pairs. *Biochemistry* **2019**, *58*, 2474–2487. [CrossRef]
58. Fernández, I.S.; Ng, C.L.; Kelley, A.C.; Wu, G.; Yu, Y.T.; Ramakrishnan, V. Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature* **2013**, *500*, 107–110. [CrossRef]
59. Ogle, J.M.; Ramakrishnan, V. Structural Insights into translational fidelity. *Annu. Rev. Biochem.* **2005**, *74*, 129–177. [CrossRef]
60. Chen, J.L.; Dishler, A.L.; Kennedy, S.D.; Yildirim, I.; Liu, B.; Turner, D.H.; Serra, M.J. Testing the nearest neighbor model for canonical RNA base pairs: Revision of GU parameters. *Biochemistry* **2012**, *51*, 3508–3522. [CrossRef]
61. Wohlgemuth, I.; Pohl, C.; Mittelstaet, J.; Konevega, A.L.; Rodnina, M.V. Evolutionary optimization of speed and accuracy of decoding on the ribosome. *Philos. Trans. R. Soc. B Biol. Sci.* **2011**, *366*, 2979–2986. [CrossRef]
62. Wohlgemuth, I.; Pohl, C.; Rodnina, M.V. Optimization of speed and accuracy of decoding in translation. *EMBO J.* **2010**, *29*, 3701–3709. [CrossRef]
63. Rozov, A.; Demeshkina, N.; Westhof, E.; Yusupov, M.; Yusupova, G. New Structural Insights into Translational Miscoding. *Trends Biochem. Sci.* **2016**, *41*, 798–814. [CrossRef]
64. Ogle, J.M.; Brodersen, D.E.; Clemons, W.M.; Tarry, M.J.; Carter, A.P.; Ramakrishnan, V. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **2001**, *292*, 897–902. [CrossRef]
65. Jenner, L.B.; Demeshkina, N.; Yusupova, G.; Yusupov, M. Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat. Struct. Mol. Biol.* **2010**, *17*, 555–560. [CrossRef]
66. Selmer, M.; Dunham, C.M.; Murphy, F.V., 4th; Weixlbaumer, A.; Petry, S.; Kelley, A.C.; Weir, J.R.; Ramakrishnan, V. Structure of the 70S Ribosome Complexed with mRNA and tRNA. *Science* **2006**, *313*, 1935–1942. [CrossRef]
67. Demeshkina, N.; Jenner, L.; Yusupova, G.; Yusupov, M. Interactions of the ribosome with mRNA and tRNA. *Curr. Opin. Struct. Biol.* **2010**, *20*, 325–332. [CrossRef]

68. Murgola, E.J.; Hijazi, K.A.; Goringer, H.U.; Dahlberg, A.E. Mutant 16S ribosomal RNA: A codon-specific translational suppressor. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 4162–4165. [CrossRef]
69. Prescott, C.D.; Dahlberg, A.E. A single base change at 726 in 16S rRNA radically alters the pattern of proteins synthesized in vivo. *EMBO J.* **1990**, *9*, 289–294. [CrossRef]
70. Moine, H.; Dahlberg, A.E. Mutations in Helix 34 of Escherichia coli 16 S Ribosomal RNA Have Multiple Effects on Ribosome Function and Synthesis. *J. Mol. Biol.* **1994**, *243*, 402–412. [CrossRef]
71. Ledoux, S.; Uhlenbeck, O.C. Different aa-tRNAs are selected uniformly on the ribosome. *Mol. Cell* **2008**, *31*, 114–123. [CrossRef]
72. Gross, L.; Schaeffer, L.; Alghoul, F.; Hayek, H.; Allmang, C.; Eriani, G.; Martin, F. Tracking the m 7 G-cap during translation initiation by crosslinking methods. *Methods* **2018**, *137*, 3–10. [CrossRef]
73. Martin, F.; Ménétret, J.F.; Simonetti, A.; Myasnikov, A.G.; Vicens, Q.; Prongidi-Fix, L.; Natchiar, S.K.; Klaholz, B.P.; Eriani, G. Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. *Nat. Commun.* **2016**, *7*, 12622. [CrossRef]
74. Martin, F.; Barends, S.; Jaeger, S.; Schaeffer, L.; Prongidi-Fix, L.; Eriani, G. Cap-assisted internal initiation of translation of histone h4. *Mol. Cell* **2011**, *41*, 197–209. [CrossRef]
75. Saint-Léger, A.; Bello, C.; Dans, P.D.; Torres, A.G.; Novoa, E.M.; Camacho, N.; Orozco, M.; Kondrashov, F.A.; Ribas de Pouplana, L. Saturation of recognition elements blocks evolution of new tRNA identities. *Sci. Adv.* **2016**, *2*, e1501860. [CrossRef]
76. Chang, A.T.; Nikonowicz, E.P. Solution NMR determination of hydrogen bonding and base pairing between the glyQS T box riboswitch Specifier domain and the anticodon loop of tRNA(Gly). *FEBS Lett.* **2013**, *587*, 3495–3499. [CrossRef]
77. Grundy, F.J.; Winkler, W.C.; Henkin, T.M. tRNA-mediated transcription antitermination in vitro: Codon-anticodon pairing independent of the ribosome. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11121–11126. [CrossRef]
78. Niyomporn, B.; Dahl, J.L.; Strominger, J.L. Biosynthesis of the peptidoglycan of bacterial cell walls. IX. Purification and properties of glycyl transfer ribonucleic acid synthetase from Staphylococcus aureus. *J. Biol. Chem.* **1968**, *243*, 773–778.




© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).